GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 18:07:07 ; Search time 251 Seconds
(without alignments)
10371.753 Million cell updates/sec

Title: US-09-805-694B-1
Perfect score: 1156
Sequence: 1 gcggccgcatgggtttcctt.....tctcctctttgagcggccgc 1156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1156	100.0	1156	22	AAD17527	Soybean vacuolar p
	2	1156	100.0	2970	22	AAD17528	Kunitz soybean try
	3	167	14.4	1807	18	AAT74041	Soybean thiol prot
	4	154.4	13.4	1056	18	AAT74043	Soybean thiol prot
	5	154.4	13.4	1392	24	ABN83213	Soybean D3-beta cD
	6	154.4	13.4	1849	18	AAT74042	Soybean thiol prot
	7	152.8	13.2	1056	18	AAT79546	Glycine max aminop
	8	151.2	13.1	1577	18	AAT90169	Oil seed rape cyst
	9	147	12.7	1553	18	AAT90170	Oil seed rape cyst
	10	136.6	11.8	1083	21	AAC44888	Arabidopsis thalia
	11	135.4	11.7	1538	21	AAC46206	Arabidopsis thalia
	12	134.2	11.6	1062	22	AAF44804	Cysteine protease
	13	132.8	11.5	1600	22	AAF44793	Cysteine protease
	14	130.8	11.3	1698	21	AAC40001	Arabidopsis thalia
	15	128.6	11.1	1369	21	AAC51084	Arabidopsis thalia
	16	128.6	11.1	2167	22	AAF44797	Cysteine protease
	17	126	10.9	1160	21	AAC33767	Arabidopsis thalia
	18	125	10.8	1201	21	AAC41593	Arabidopsis thalia
	19	121.8	10.5	1112	21	AAC42369	Arabidopsis thalia
	20	120.2	10.4	1434	18	AAT90172	Oil seed rape cyst
	21	120	10.4	910	19	AAV66759	Pathogen response
	22	118.6	10.3	1390	18	AAT90171	Oil seed rape cyst
	23	118.6	10.3	1441	18	AAT90173	Oil seed rape cyst
	24	117.2	10.1	765	22	AAF44805	Cysteine protease
	25	117	10.1	1186	16	AAQ94588	B. campestris cDNA
	26	117	10.1	1186	18	AAT65477	EA9 DNA. Brassica
	27	117	10.1	1186	19	AAV36970	Brassica campestri
	28	117	10.1	1186	21	AAZ35693	Brassica campestri
	29	117	10.1	1186	22	AAH47380	B. campestris seed
	30	115.4	10.0	1186	19	AAV27947	Brassica campestri
	31	114.8	9.9	1474	18	AAT90174	Oil seed rape cyst
	32	113	9.8	1370	21	AAC32968	Arabidopsis thalia
	33	109	9.4	1645	20	AAV82458	Triticum sp. cyste
	34	108.2	9.4	1879	21	AAC44552	Zea mays DNA fragm
	35	103.8	9.0	2446	21	AAC46450	Arabidopsis thalia
	36	99	8.6	1224	21	AAC40331	Arabidopsis thalia
	37	98.8	8.5	1474	22	AAF44796	Cysteine protease
	38	97.6	8.4	883	22	AAF44791	Cysteine protease
	39	97.6	8.4	948	22	AAF44790	Cysteine protease
	40	95.8	8.3	1077	21	AAC51435	Arabidopsis thalia
c	41	95.8	8.3	1282	24	ABN98252	Arabidopsis thalia
	42	93.2	8.1	1305	21	AAC47882	Arabidopsis thalia
	43	92.6	8.0	1407	22	AAF44802	Cysteine protease
	44	92.6	8.0	1463	22	AAF44787	Cysteine protease
	45	92.4	8.0	785	21	AAC33494	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAD17527
ID AAD17527 standard; DNA; 1156 BP.
XX
AC AAD17527;
XX
DT 10-DEC-2001 (first entry)
XX
DE Soybean vacuolar protein (P34) encoding DNA.
XX
KW Soybean; allergen; transgenic plant; P34 protein; Gly m Bd 30K; Gly m 1;
KW soybean vacuolar protein; Gly m IA; Gly m IB; rGLY m3; Glycinin G1;
KW alaB1b; food; infant formula; animal feed; coating; salad oil; syrup;
KW spraying oil; roasting oil; frying oil; cracker; confectionery product;
KW snack food; topping; sauce; batter; breading mixture; baking mix; dough;
KW ds.
XX
OS Glycine max.
XX
PN WO200168887-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-US08254.
XX
PR 16-MAR-2000; 2000US-0189823.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Jung R, Kinney AJ;
XX

```
DR   WPI; 2001-582460/65.
XX
PT   Recombinant expression construct to lower allergen (e.g. Gly m Bd 30K)
PT   content of a soybean, comprises a nucleic acid fragment encoding the
PT   allergen, useful for producing soybean plants which can be used to make
PT   soybean products
XX
PS   Claim 1; Page 48; 57pp; English.
XX
CC   The patent discloses hypoallergenic transgenic soybeans and recombinant
CC   expression constructs to lower soybean vacuolar protein, commonly known
CC   as P34 (Gly m Bd 30K or Gly m 1) and other allergens such as Gly m IA,
CC   Gly m IB, rGLY m3 or Glycinin G1 (alaB1b). The allergen content of the
CC   soybean is reduced by sense suppression which is accomplished by using
CC   the expression construct that comprises a nucleic acid fragment encoding
CC   the allergen. The constructs are useful for producing hypoallergenic
CC   transgenic soybean plants which can be used to make hypoallergenic
CC   soybean products which can be used in a variety of food (e.g. infant
CC   formulas) and animal feed applications. The oil made from seeds of the
CC   hypoallergenic transgenic soybean plants can be used as ingredients,
CC   as coatings, as salad oils, as spraying oils, as roasting oils, and
CC   as frying oils. The foods in which the oil may be used include crackers
CC   and snack foods, confectionery products, syrups and toppings, sauces,
CC   batter and breading mixtures, baking mixes and doughs. The present
CC   sequence is a DNA encoding soybean vacuolar protein, P34.
XX
SQ   Sequence 1156 BP; 361 A; 223 C; 277 G; 295 T; 0 other;

  Query Match             100.0%;  Score 1156;  DB 22;  Length 1156;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1156;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 60

Qy     61 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 120

Qy    121 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 180

Qy    181 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 240

Qy    241 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 300

Qy    301 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 360

Qy    361 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 420

Qy    421 CATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 480

Qy    481 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 540

Qy    541 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 600

Qy    601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660

Qy    661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 720

Qy    721 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 780

Qy    781 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 840

Qy    841 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 900

Qy    901 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 960

Qy    961 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1020

Qy   1021 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1080

Qy   1081 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1140

Qy   1141 CTCTTTGAGCGGCCGC 1156
          ||||||||||||||||
Db   1141 CTCTTTGAGCGGCCGC 1156


RESULT 2
AAD17528
ID   AAD17528 standard; DNA; 2970 BP.
XX
AC   AAD17528;
XX
DT   10-DEC-2001  (first entry)
XX
DE   Kunitz soybean trypsin inhibitor (KSTI) encoding DNA.
XX
KW   Soybean; allergen; transgenic plant; P34 protein; Gly m Bd 30K; Gly m 1;
KW   soybean vacuolar protein; Gly m IA; Gly m IB; rGLY m3; Glycinin G1;
KW   alaB1b; food; infant formula; animal feed; coating; salad oil; syrup;
KW   spraying oil; roasting oil; frying oil; cracker; confectionery product;
KW   snack food; topping; sauce; batter; breading mixture; baking mix; dough;
KW   Kunitz soybean trypsin inhibitor; KSTI; KTi3; ds.
XX
OS   Glycine max.
XX
PN   WO200168887-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-US08254.
XX
PR   16-MAR-2000; 2000US-0189823.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Jung R,  Kinney AJ;
XX
DR   WPI; 2001-582460/65.
```

```
XX
PT   Recombinant expression construct to lower allergen (e.g. Gly m Bd 30K)
PT   content of a soybean, comprises a nucleic acid fragment encoding the
PT   allergen, useful for producing soybean plants which can be used to make
PT   soybean products
XX
PS   Claim 4; Page 48-49; 57pp; English.
XX
CC   The patent discloses hypoallergenic transgenic soybeans and recombinant
CC   expression constructs to lower soybean vacuolar protein, commonly known
CC   as P34 (Gly m Bd 30K or Gly m 1) and other allergens such as Gly m IA,
CC   Gly m IB, rGLY m3 or Glycinin G1 (alaB1b). The allergen content of the
CC   soybean is reduced by sense suppression which is accomplished by using
CC   the expression construct that comprises a nucleic acid fragment encoding
CC   the allergen. The constructs are useful for producing hypoallergenic
CC   transgenic soybean plants which can be used to make hypoallergenic
CC   soybean products which can be used in a variety of food (e.g. infant
CC   formulas) and animal feed applications. The oil made from seeds of the
CC   hypoallergenic transgenic soybean plants can be used as ingredients,
CC   as coatings, as salad oils, as spraying oils, as roasting oils, and
CC   as frying oils. The foods in which the oil may be used include crackers
CC   and snack foods, confectionery products, syrups and toppings, sauces,
CC   batter and breading mixtures, baking mixes and doughs. The present
CC   sequence is a DNA encoding kunitz soybean trypsin inhibitor (KSTI
CC   or KTi3), a minor soybean seed allergen.
XX
SQ   Sequence 2970 BP; 1018 A; 547 C; 527 G; 878 T; 0 other;

  Query Match             100.0%;  Score 1156;  DB 22;  Length 2970;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1156;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     622 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 681

Qy      61 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     682 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 741

Qy     121 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     742 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 801

Qy     181 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     802 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 861

Qy     241 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     862 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 921

Qy     301 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     922 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 981

Qy     361 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     982 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 1041

Qy     421 CATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1042 CATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 1101

Qy     481 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1102 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 1161

Qy     541 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1162 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 1221

Qy     601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1222 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 1281

Qy     661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1282 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 1341

Qy     721 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1342 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 1401

Qy     781 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1402 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 1461

Qy     841 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1462 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 1521

Qy     901 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1522 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 1581

Qy     961 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1582 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1641

Qy    1021 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1642 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1701

Qy    1081 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1702 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1761

Qy    1141 CTCTTTGAGCGGCCGC 1156
          ||||||||||||||||
Db    1762 CTCTTTGAGCGGCCGC 1777
```

JP497

```
RESULT 3
AAT74041
ID   AAT74041 standard; cDNA to mRNA; 1807 BP.
XX
AC   AAT74041;
XX
DT   17-SEP-1997  (first entry)
XX
DE   Soybean thiol protease D3-alpha encoding cDNA.
XX
KW   Soybean; cotyledon; germinate; transformant; Escherichia coli;
KW   recombinant expression vector; ds.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             68..1459
FT                   /*tag=  a
FT                   /product= Thiol_protease
FT   sig_peptide     68..460
FT                   /*tag=  b
FT   mat_peptide     461..1459
FT                   /*tag=  c
FT                   /product= D3-alpha
XX
PN   JP09121870-A.
XX
PD   13-MAY-1997.
XX
PF   28-DEC-1995;   95JP-0353931.
XX
PR   30-AUG-1995;   95JP-0245279.
PR   29-DEC-1994;   94JP-0340399.
```

JP09121870A

```
XX
PA   (AJIN ) AJINOMOTO KK.
XX
DR   WPI; 1997-314230/29.
DR   P-PSDB; AAW19541.
XX
PT   New thiol protease and corresponding DNA sequence - used in the
PT   preparation of the thiol protease
XX
PS   Claim 1; Page 10-12; 17pp; Japanese.
XX
CC   The present sequence encodes a novel thiol protease D3-alpha derived
CC   from germinated soybean cotyledon. A method for the production of
CC   thiol protease using Escherichia coli transformed with recombinant
CC   expression vectors containing a DNA sequence encoding thiol protease has
CC   been produced. The method can be used to prepare large amounts of thiol
CC   protease with high efficiency.
XX
SQ   Sequence 1807 BP; 500 A; 387 C; 449 G; 471 T; 0 other;

  Query Match             14.4%;  Score 167;  DB 18;  Length 1807;
  Best Local Similarity   53.0%;  Pred. No. 6.4e-39;
  Matches  493;  Conservative    0;  Mismatches  405;  Indels   33;  Gaps    5;

Qy     103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
           ||| ||    ||   |  |  || ||   ||  ||| | | |  |||  |    |||| |
Db     168 CCACGTCGCGCAGCGACGAGGAGCTGATGTCCATGTACGAGCAGTGGCTGGTGAAGCACG 227

Qy     163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
           |    || ||| |    | ||  ||  |||  ||||| |  |||| |||||| | |||
Db     228 GCAAGGTGTACAACGCGCTCGGGGAGAAGGAGAAGAGGTTCCAGATCTTCAAGGACAACC 287

Qy     223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
           |   |  |||   |||   ||  |  |      |   | || |     |      | |||
Db     288 TGCGATTCATCGACGACCACAACTCCCAAGAGGACCGAACCTACAAGC------TCGGAC 341

Qy     283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
           |||||  ||| || ||  ||||     | |||| |||    |||||||||  | |    |
Db     342 TGAACCGGTTCGCGGATCTCACCAACGAGGAGTACAGGGCCAAGTACTTGGGAACCAAGA 401

Qy     343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
             |||      | |    ||   | | ||   |  ||     |   |  ||    |
Db     402 TCGATCCCAACCGGAGGCTCGGCAAGACCCCGAGCAACCGATACGCGCCACG---TGTCG 458

Qy     403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
           | ||| | | ||||| |||    ||||||||||  ||||||   |  | | ||| ||
Db     459 GCGACAAACTACCTGAATCGGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCAGTCAAAG 518

Qy     463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
           ||||||| |||||||| ||  |||||||| || || || |  || ||  | |||| |  |
Db     519 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTGGAAGGAATAA 578

Qy     523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
           ||   ||||  ||||| || ||  ||    | || |||||||| | || || ||||
Db     579 ATAAGATAGTGACAGGTGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 638

Qy     580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
             | | | | |||||| |    |    |||     | |  |  | || || |   || |
Db     639 CAGGATATAACGAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATTATCA 698

Qy     640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
           |  |||| |||||||   |||| || ||||| |||||| | | |   || |||||||||
Db     699 ACAATGGCGGCATTGATTCTGAAGAGGATTACCCTTACCGTGGTGTTGATGGTAGATGTG 758

Qy     700 AAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAG 759
           |  |  ||| || | | |  ||||||     |   |   |  |        |     |
Db     759 ACACATATAGGAAAAATGCTAAGGTTGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 818

Qy     760 ATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAA 819
           ||||                 |||  ||| ||   ||  ||  |     | || ||
Db     819 ATGAT-----------------GAGTTAGCTTTGAAAAAGGCTGTTGCAAATCAGCCCG 860

Qy     820 TTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAG 879
```

```
           | |||||  | ||||| | |   |      | ||  |       |   | | || || |
Db     861 TCAGTGTAGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTGTAT 917

Qy     880 AAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAG 939
           |   | | | ||      |   |  ||||  |||  |  ||| || |||||| |||
Db     918 TCACTGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGTACAG 977

Qy     940 CGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATG 999
           |  |||||   ||||| ||||| |  |  ||||||||||       ||||||||||||
Db     978 CTAATGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTCCTAGCTGGGGAGAAGATG 1037

Qy    1000 GTTACATTTGGATCCAAAGAAACACGGGTAA 1030
           | |||||  || |  |||||||    | |||
Db    1038 GCTACATCAGGTTAGAAAGAAATCTTGCTAA 1068


RESULT 4
AAT74043
ID   AAT74043 standard; cDNA to mRNA; 1056 BP.
XX
AC   AAT74043;
XX
DT   17-SEP-1997  (first entry)
XX
DE   Soybean thiol protease D3-beta encoding partial cDNA fragment.
XX
KW   Soybean; cotyledon; germinate; transformant; Escherichia coli;
KW   recombinant expression vector; ds.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   iDNA            1..9
FT                   /*tag=  a
FT                   /label= Insertion_sequence
FT   CDS             1..1056
FT                   /*tag=  b
FT                   /product= Thiol_protease_D3-beta
FT                   /note= "Partial fragment"
XX
PN   JP09121870-A.
XX
PD   13-MAY-1997.
XX
PF   28-DEC-1995;   95JP-0353931.
XX
PR   30-AUG-1995;   95JP-0245279.
PR   29-DEC-1994;   94JP-0340399.
XX
PA   (AJIN ) AJINOMOTO KK.
XX
DR   WPI; 1997-314230/29.
DR   P-PSDB; AAW23504.
XX
PT   New thiol protease and corresponding DNA sequence - used in the
PT   preparation of the thiol protease
XX
PS   Claim 5; Page 14-15; 17pp; Japanese.
XX
CC   The present sequence represents a partial cDNA fragment encoding
CC   the N-terminal amino acids of the novel thiol protease D3-beta,
CC   derived from germinated soybean cotyledon. A method for the production
CC   of thiol protease using Escherichia coli transformed with recombinant
CC   expression vectors containing the present sequence has been produced
CC   (plasmid pGEMpbeta). The method can be used to prepare large amounts
CC   of thiol protease D3-beta with high efficiency.
XX
SQ   Sequence 1056 BP; 295 A; 229 C; 297 G; 235 T; 0 other;

  Query Match             13.4%;  Score 154.4;  DB 18;  Length 1056;
  Best Local Similarity   52.9%;  Pred. No. 2.7e-35;
  Matches  494;  Conservative    0;  Mismatches  401;  Indels   39;  Gaps    6;
```

```
Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
Db     38 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 97

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
Db     98 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 157

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
Db    158 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 211

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
Db    212 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 271

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
Db    272 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 328

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
Db    329 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 388

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
Db    389 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 448

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
Db    449 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 508

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
Db    509 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 568

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
Db    569 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 628

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756
Db    629 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 688

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
Db    689 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 727

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
Db    728 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 784

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
Db    785 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 844

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
Db    845 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 904

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
Db    905 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 938


RESULT 5
ABN83213
ID   ABN83213 standard; cDNA; 1392 BP.
XX
AC   ABN83213;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Soybean D3-beta cDNA.
XX
KW   Soybean; angiotensin converting enzyme inhibitor; hypertension;
KW   hypotensive; taste; D3-beta; gene; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1392
FT                   /*tag= a
FT                   /product= "D3-beta"
FT                   /partial
FT                   /note= "No stop codon given"
FT   sig_peptide     1..396
FT                   /*tag= b
FT   mat_peptide     397..1392
FT                   /*tag= c
FT                   /note= "Mature D3-beta"
XX
PN   WO200255546-A1.
XX
PD   18-JUL-2002.
XX
PF   15-JAN-2002; 2002WO-JP00194.
XX
PR   16-JAN-2001; 2001JP-0007400.
PR   04-OCT-2001; 2001JP-0308974.
XX
PA   (AJIN ) AJINOMOTO CO INC.
XX
PI   Kodera T,  Nio N;
XX
DR   WPI; 2002-520117/55.
DR   P-PSDB; ABB81808.
XX
PT   Peptides, useful as hypotensive agents or in health foods -
XX
PS   Example 1; Page 30-34; 43pp; Japanese.
XX
CC   The invention relates to a novel set of peptides and their salts, which
CC   act as inhibitors of angiotensin converting enzyme. The peptides of the
CC   invention have hypotensive activity. The peptides are used as hypotensive
CC   agents or in health foods, and have favourable taste. The present
CC   sequence encodes a soybean D3-beta protein, used in example 1 of the
CC   invention.
XX
SQ   Sequence 1392 BP; 364 A; 319 C; 379 G; 330 T; 0 other;

  Query Match           13.4%;  Score 154.4;  DB 24;  Length 1392;
  Best Local Similarity  52.9%;  Pred. No. 3.1e-35;
  Matches  494;  Conservative    0;  Mismatches  401;  Indels   39;  Gaps    6;

Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
Db    104 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 163

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
Db    164 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 223

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
Db    224 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 277

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
Db    278 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 337

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
Db    338 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 394

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
```

```
Db    395 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 454

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522

Db    455 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 514

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579

Db    515 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 574

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639

Db    575 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 634

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699

Db    635 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 694

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756

Db    695 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 754

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816

Db    755 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 793

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876

Db    794 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 850

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936

Db    851 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 910

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996

Db    911 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 970

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030

Db    971 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 1004
```

RESULT 6
AAT74042
ID AAT74042 standard; cDNA to mRNA; 1849 BP.
XX
AC AAT74042;
XX
DT 17-SEP-1997 (first entry)
XX
DE Soybean thiol protease D3-beta encoding cDNA.
XX
KW Soybean; cotyledon; germinate; transformant; Escherichia coli;
KW recombinant expression vector; ds.
XX
OS Glycine max.
XX
FH Key   Location/Qualifiers
FT CDS   48..1442
FT   /*tag= a
FT   /product= Thiol_protease
FT sig_peptide   48..443
FT   /*tag= b
FT mat_peptide   444..1442
FT   /*tag= c
FT   /product= D3-beta
XX
PN JP09121870-A.
XX
PD 13-MAY-1997.
XX
PF 28-DEC-1995; 95JP-0353931.
XX
PR 30-AUG-1995; 95JP-0245279.
PR 29-DEC-1994; 94JP-0340399.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR WPI; 1997-314230/29.
DR P-PSDB; AAW19542.
XX
PT New thiol protease and corresponding DNA sequence - used in the
PT preparation of the thiol protease
XX
PS Claim 13; Page 12-14; 17pp; Japanese.
XX
CC The present sequence encodes a novel thiol protease D3-beta derived
CC from germinated soybean cotyledon. A method for the production of
CC thiol protease using Escherichia coli transformed with recombinant
CC expression vectors containing a DNA sequence encoding thiol protease has
CC been produced. The method can be used to prepare large amounts of thiol
CC protease with high efficiency.
XX
SQ Sequence 1849 BP; 539 A; 394 C; 454 G; 462 T; 0 other;

Query Match 13.4%; Score 154.4; DB 18; Length 1849;
Best Local Similarity 52.9%; Pred. No. 3.6e-35;
Matches 494; Conservative 0; Mismatches 401; Indels 39; Gaps 6;

```
Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162

Db    151 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 210

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222

Db    211 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 270

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282

Db    271 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 324

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342

Db    325 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 384

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402

Db    385 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 441

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462

Db    442 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 501

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522

Db    502 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 561

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579

Db    562 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 621

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639

Db    622 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 681

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699

Db    682 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 741

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756

Db    742 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 801

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
```

```
          ||||||                        ||| ||   ||  ||| |     | || |
Db    802 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 840

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
          |  | || ||  | ||||| | |    |      | ||  |        |   | | || ||
Db    841 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 897

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
          |    |   | | | ||      |   |  ||||   |||  |   ||| || |||||
Db    898 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 957

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
          ||||  | |||   ||||| ||||| |  |  ||||||||||||      |||||||| |
Db    958 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 1017

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
          |||| |||||  |  |  |||||||    | |||
Db   1018 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 1051
```

RESULT 7
AAT79546
ID AAT79546 standard; cDNA to mRNA; 1056 BP.
XX
AC AAT79546;
XX
DT 09-JAN-1998 (first entry)
XX
DE Glycine max aminopeptidase GX cDNA.
XX
KW Aminopeptidase; protein hydrolysis; glutamic acid; aspartic acid;
KW food industry; soy sauce; soybean; soya; ds.
XX
OS Glycine max.
XX
FH Key   Location/Qualifiers
FT CDS   1..1056
FT   /product= aminopeptidase_GX
XX
PN EP794253-A2.
XX
PD 10-SEP-1997.
XX
PF 07-MAR-1997; 97EP-0103862.
XX
PR 14-FEB-1997; 97JP-0030458.
PR 08-MAR-1996; 96JP-0051848.
XX
PA (AJIN) AJINOMOTO KK.
XX
PI Asano M, Kawai M, Miwa T, Nio N;
XX
DR WPI; 1997-437478/41.
DR P-PSDB; AAW23147.
XX
PT Aminopeptidase GX - for hydrolysing peptide(s) and proteins having
PT amino-terminal glutamic or aspartic acid residue
XX
PS Disclosure; Pages 22-24; 45pp; English.
XX
CC This cDNA encodes a novel aminopeptidase GX which acts on a protein or
CC oligopeptide having glutamic acid or aspartic acid at the N terminal.
CC This aminopeptidase will act on peptides hardly decomposed with
CC exopeptidases (including other known aminopeptidases and
CC carboxypeptidases). The enzyme, or a cell extract containing it
CC was obtained by disrupting germinated soya cotyledons. The resulting
CC protein hydrolysates can be used in foods, seasonings and feeds.
XX
SQ Sequence 1056 BP; 295 A; 230 C; 297 G; 234 T; 0 other;

  Query Match   13.2%; Score 152.8; DB 18; Length 1056;
  Best Local Similarity 52.8%; Pred. No. 8.2e-35;
  Matches 493; Conservative   0; Mismatches 402; Indels   39; Gaps   6;

```
Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
          ||| |||   |||  || |  || ||   ||  ||| | | |  |||       |||| |
Db     38 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 97

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
          |    || ||| |    | ||  ||  |||  ||| |  |  |||| |||||| | |||
Db     98 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 157

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
          |   |  |||   |||  ||  |         |   | || |    |       | |||
Db    158 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 211

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
          |||||  ||| |||||  ||||     | || | |||    |||||||||  | |    |
Db    212 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 271

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
          |||       | |    ||  || | ||   |  ||     |   |  ||   |
Db    272 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 328

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
          | ||| |    |||| ||    ||||||||||  ||||||   |  | |  || ||
Db    329 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 388

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
          |||||||  |||||||| ||  |||||||  || || || |  || ||  |||||| |  |
Db    389 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 448

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
          ||  |||| ||||||  || ||  ||    | || |||||||||  | || || ||||
Db    449 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 508

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
          | | | | | |||| |   |    |||      | |   |  | || || |    | |
Db    509 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 568

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
          |  |||| |||||||   |||||||| ||||| || ||| | | |   || |||||||||
Db    569 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 628

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756
          |  |  ||| || | | |  ||    ||| | ||||| |  || |||  | |       |
Db    629 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 688

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
          ||||||                          ||| ||   ||  ||| |     | || |
Db    689 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 727

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
          |  | || ||  | ||||| | |    |      | ||  |        |   | | || ||
Db    728 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 784

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
          |    |   | | | ||      |   |  ||   |||  |   ||| || |||||
Db    785 TATTCACGGGGAGATGTGGCACAGCACTAGACCATGGTGTCGTGGCTGTTGGGTATGGAA 844

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
          ||||  | |||   ||||| ||||| |  |  ||||||||||       |||||||| |
Db    845 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 904

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
          |||| |||||  |  |  |||||||    | |||
Db    905 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 938
```

RESULT 8
AAT90169
ID AAT90169 standard; cDNA; 1577 BP.
XX
AC AAT90169;
XX
DT 20-APR-1998 (first entry)

```
XX
DE   Oil seed rape cysteine protease cDNA clone CDCYS12.
XX
KW   Promoter; oil seed rape; cDNA clone CDCYS12;
KW   cysteine protease; disrupter protein; plant; tolerance; herbicide;
KW   insect pest; fungal disease; improved yield; improved quality;
KW   fertility control; ss.
XX
OS   Brassica napus.
XX
FH   Key             Location/Qualifiers
FT   CDS             30..1361
FT                   /transl_except= (pos:690..692, aa:Xaa)
FT                   /note= "Xaa = unknown"
FT                   /*tag=  a
XX
PN   WO9735983-A2.
XX
PD   02-OCT-1997.
XX
PF   18-MAR-1997;   97WO-GB00729.
XX
PR   22-MAR-1996;   96GB-0006062.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Greenland AJ,  Jepson I,  Thomas DRP;
XX
DR   WPI; 1997-489646/45.
DR   P-PSDB; AAW27441.
XX
PT   Cysteine protease promoter - obtained from oil seed rape, useful to
PT   develop plants with improved agronomic characteristics
XX
PS   Claim 18; Fig 12; 137pp; English.
XX
CC   The present sequence encodes an oil seed rape cysteine protease,
CC   the promoter for which can restrict the expression of a disrupter
CC   protein gene to a suitable stage of plant development, to provide
CC   plants with novel agronomic features, e.g. tolerance to herbicides,
CC   insect pests and fungal diseases, improved yield and/or quality of
CC   harvested product, and novel mechanisms for the control of plant
CC   fertility.
XX
SQ   Sequence 1577 BP; 458 A; 325 C; 356 G; 437 T; 1 other;

  Query Match             13.1%;  Score 151.2;  DB 18;  Length 1577;
  Best Local Similarity   55.2%;  Pred. No. 3e-34;
  Matches  369;  Conservative    1;  Mismatches  274;  Indels   24;  Gaps    3;

Qy     414 CCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGC 473
           || |  ||    |||||||||||| ||||| |   | ||  | || ||  | |||||  ||
Db     402 CCAGATTCTGTTGATTGGAGGAAGAAAGGAGCTGTTACTAATGTCAAAGATCAAGGAAGC 461

Qy     474 TGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCA 533
           || |||  | ||||| | ||| || || || ||||| || |||| |    |    || | |
Db     462 TGCGGAGCGTGTTGGTCTTTCTCGGCGACTGGAGCTATGGAAGGAATCAACCAGATTGTA 521

Qy     534 ACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAA 593
           |||||||| ||  | ||||| ||||| |||||||||| | || ||||   |   |  | |
Db     522 ACAGGAGATCTCATCAGCCTCTCTGAGCAAGAACTCATTGATTGTGATAAGTCATACAAC 581

Qy     594 GGTTCTTACAATGGATGGC---AGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGC 650
           | |   | ||||||  | |    | |  |  | ||  |||  ||  |  || |   |||
Db     582 GATGGATGCAATGGTGGTCTCATGGACTACGCTTTTCAATTTGTCATTAAAAACCATGGG 641

Qy     651 ATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAG 710
           |||| ||| ||  | ||||||||||   ||     || || |  || :|     |||||
Db     642 ATTGACACAGAGAAAGATTATCCTTATCAAGAACGTGATGGCACCTGTMAGAAAGATAAG 701

Qy     711 ATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACA 770
             |  |    ||||||   |     |      |  ||    |||  ||| ||||    | |
Db     702 TTGAATAGAAAGGTTGTGACAATTGATAGCTACGCTGGTGTAAAATCAAATGACGAGAAA 761

Qy     771 GAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCA 830
           |  | |     ||    ||||       |  |   |||           || |||
Db     762 GCGTTACTAGAAGCTGTAGCGGCTCAGCCAGTTAGTGTTG-----------GTATCTGT 809

Qy     831 ATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACA 890
             |  | ||  ||||| |||||| |   ||||||  |    |||   |   |  |||
Db     810 GGGAGCGAGAGAGCGTTTCAGTTATACTCTAAGGGAATATTCTCTGGCCCATGTTCAACA 869

Qy     891 AGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTA 950
             |          | |||   ||  |  | || || || ||||||  | | |||||
Db     870 TCATTG---------GACCACGCAGTGCTCATCGTAGGATACGGTTCAAAGAACGGTGTT 920

Qy     951 GATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGG 1010
           ||||||||||| | ||| || || |||||| |   |||||||   |||||  |  ||
Db     921 GATTACTGGATCGTGAAGAACTCTTGGGGAAAGAGTTGGGGAATGGATGGGTTTATCCAC 980

Qy    1011 ATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATAC 1070
           || ||  | ||||| || ||   |  |||||| || || || ||     | ||||||||
Db     981 ATGCAGCGTAACACCGGCAACGCAGAAGGAGTATGCGGAATCAACATGCTGGCTTCATAT 1040

Qy    1071 CCAACCAA 1078
           || | |||
Db    1041 CCCATCAA 1048


RESULT 9
AAT90170
ID   AAT90170 standard; cDNA; 1553 BP.
XX
AC   AAT90170;
XX
DT   20-APR-1998  (first entry)
XX
DE   Oil seed rape cysteine protease cDNA clone CDCYS14.
XX
KW   Promoter; oil seed rape; cDNA clone CDCYS14;
KW   cysteine protease; disrupter protein; plant; tolerance; herbicide;
KW   insect pest; fungal disease; improved yield; improved quality;
KW   fertility control; ss.
XX
OS   Brassica napus.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1369
FT                   /*tag=  a
XX
PN   WO9735983-A2.
XX
PD   02-OCT-1997.
XX
PF   18-MAR-1997;   97WO-GB00729.
XX
PR   22-MAR-1996;   96GB-0006062.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Greenland AJ,  Jepson I,  Thomas DRP;
XX
DR   WPI; 1997-489646/45.
DR   P-PSDB; AAW27442.
XX
PT   Cysteine protease promoter - obtained from oil seed rape, useful to
PT   develop plants with improved agronomic characteristics
XX
PS   Claim 19; Fig 13; 137pp; English.
XX
CC   The present sequence encodes an oil seed rape cysteine protease,
CC   the promoter for which can restrict the expression of a disrupter
CC   protein gene to a suitable stage of plant development, to provide
CC   plants with novel agronomic features, e.g. tolerance to herbicides,
CC   insect pests and fungal diseases, improved yield and/or quality of
CC   harvested product, and novel mechanisms for the control of plant
```

```
CC   fertility.
XX
SQ   Sequence 1553 BP; 432 A; 329 C; 360 G; 432 T; 0 other;

  Query Match            12.7%;  Score 147;  DB 18;  Length 1553;
  Best Local Similarity  54.6%;  Pred. No. 5.2e-33;
  Matches  372;  Conservative    0;  Mismatches  285;  Indels   24;  Gaps    3;

Qy    401 TTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAA 460
          | | | | |    || |  ||    ||||||||||| ||||| |   | ||  | || ||
Db    397 TGGGGGCAAAGTTCCAGATTCTGTTGATTGGAGGAAGAAAGGAGCTGTTACTAATGTCAA 456

Qy    461 GTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGC 520
            | |||||  |||| |||  | ||||| | || || || || ||||| || |||| |
Db    457 AGATCAAGGAAGCTGCGGAGCGTGTTGGTCTTTCTCGGCGACTGGAGCAATGGAAGGAAT 516

Qy    521 ACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGT 580
            |    || | |||||||||| ||  | ||||| ||||| || |||||  | || ||||
Db    517 CAACCAGATTGTAACAGGAGATCTCATCAGCCTCTCTGAGCAGGAACTAATTGATTGTGA 576

Qy    581 GGAAGAAAGCGAAGGTTCTTACAATGGATGGC---AGTATCAATCGTTCGAATGGGTTTT 637
            |      | | | |   | ||||||  | |    | |  |  | ||  |||  ||  |
Db    577 CAAGTCCTACAACGATGGATGCAATGGTGGTCTCATGGACTACGCTTTTCAATTTGTCAT 636

Qy    638 AGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATG 697
            || |   ||| || | ||| ||  | ||||||||||||   ||     || || |  ||
Db    637 TAAAAACCATGGAATCGACACAGAGAAAGATTATCCTTATCAAGAACGTGATGGCACCTG 696

Qy    698 CAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTC 757
            |||    ||||| |  ||   ||||||   |     |      |  ||    |||  ||
Db    697 TAAAAAAGATAAGTTGAAAAGAAAGGTTGTGACAATTGATAGCTATGCTGGCGTAAAATC 756

Qy    758 AGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACC 817
          | | ||    | ||  | |  |  ||    ||||         |  |   |||
Db    757 AAACGACGAGAAAGCGTTACTGGAAGCTGTAGCGGCTCAGCCAGTTAGTGTTG------- 809

Qy    818 AATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGG 877
              |  |||         |  | ||  ||||| |||||  |   ||||||  |    ||
Db    810 -----GCATCTGTGGCAGCGAGAGAGCGTTTCAGCTATACTCTAAGGGAATATTCTCTG- 863

Qy    878 AGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTC 937
              |  || || |        |  | |||   ||  |  | || || || |||||
Db    864 --------GCCCATGTTCAACATCATTGGACCACGCAGTGCTCATCGTAGGATACGGTTC 915

Qy    938 AGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA 997
          |  | | ||||| |||||||||| | ||| || || |||||| |   ||||||    ||
Db    916 ACAGAACGGTGTTGATTACTGGATCGTGAAGAACTCTTGGGGAAAGAGTTGGGGTATGGA 975

Qy    998 TGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTA 1057
          ||| |  ||    || ||  | ||||| || || |    |||||| ||||| || |||
Db    976 TGGGTTTATGCACATGCAGCGTAACACCGGCAACTCGGAAGGAGTATGTGGAATCAATAT 1035

Qy   1058 TTTCGCTTCATACCCAACCAA 1078
            ||||||| || || | |||
Db   1036 GCTCGCTTCGTATCCCATCAA 1056


RESULT 10
AAC44888
ID   AAC44888 standard; DNA; 1083 BP.
XX
AC   AAC44888;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 44507.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
```

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.8%; Score 136.6; DB 21; Length 1083;
Best Local Similarity 50.4%; Pred. No. 5.3e-30;
Matches 405; Conservative 0; Mismatches 374; Indels 24; Gaps 2;

```
Qy   283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTAC------TTGCAAG 336
             | ||  ||||||||||  | ||   | | || ||  |    | ||||      ||  |||
Db   296 TAAATCAGTTTGCTGATTTAACCAATGACGAATTTCGTTCCATGTACACTGGTTTCAAAG 355

Qy   337 CTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAAT 396
              |  |  ||   | ||    |||| | |||     |    |    | | | |  || |
Db   356 GTGTCTCGGCATTATCTAGCCAAAGCCAAACTAAAATGTCGCCGTTTAGGTACCAAAACG 415

Qy   397 ATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAG 456
              ||||| ||    |   || |  ||    || |||||||| ||||| |   | ||||
Db   416 TTTCTTCTGGTGCTTTGCCGGTTTCTGTTGACTGGAGGAAGAAAGGAGCTGTGACCCCTA 475

Qy   457 TAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAG 516
             | ||| | |||||  |||| ||| |  ||||||||||||| ||    |  || || ||||
Db   476 TCAAGAATCAAGGCAGCTGCGGATGTTGTTGGGCGTTTTCAGCGGTTGCGGCTATTGAAG 535

Qy   517 CAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACT 576
              ||||     ||||   | |||  | ||| |     | || ||||||  | || || || |
Db   536 GAGCAACACAAATAAAGAAAGGGAAACTTATATCTTTGTCAGAACAACAGCTTGTTGATT 595

Qy   577 GTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTT 636
             | |    | |       || |   |    || |    | ||    |||| ||     |
Db   596 GCGACACAAACGATTTTGGCTGCGAAGGCGGTTTAATGGATACTGCGTTTGAGCATATAA 655

Qy   637 TAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGAT 696
              ||    ||| ||| |  | ||||     |||||||||||| ||   |||| | ||  |
Db   656 AAGCGACTGGCGGCTTGACAACTGAGTCAAATTATCCTTACAAAGGCGAAGACGCTACTT 715
```

```
Qy     697 GCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGT 756
           ||||  |||| |||                  |   |   | | | ||||   || |
Db     716 GCAATTCCAAAAAG-----------------ACCAATCCAAAAGCAACTTCTATTACAG 757

Qy     757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
              |||||  |          |    |||||||| |  | |  ||  |     | ||||
Db     758 GTTATGAGGATGTCCCGGTTAATGATGAGCAAGCACTGATGAAGGCAGTGGCACACCAAC 817

Qy     817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
           |  |||| ||   |||||| | |   | |||| |||| |   |   |       |  ||
Db     818 CGGTTAGCGTTGGAATTGAAGGAGGTGGTTTTGATTTCCAATTCTATTCGTCTGGTGTGT 877

Qy     877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
                     | | |||||     | ||    |  |    |||         |  |
Db     878 TCACTGGAGAGTGCACTACGTATCTTGATCATGCAGTAACTGCGATTGGATACGGCGAAT 937

Qy     937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
           |  |  | ||   | | || |||||    || ||||||||||||  | | |||||||||
Db     938 CTACTAACGGATCAAAGTATTGGATCATCAAGAATTCATGGGGAACAAAATGGGGAGAAA 997

Qy     997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATT 1056
            ||| || ||  |||| |||| | |    .    ||  |  |||| | |||||  |
Db     998 GTGGATATATGAGGATTCAAAAAGATGTCAAGGATAAACAAGGACTATGTGGTCTTGCCA 1057

Qy    1057 ATTTCGCTTCATACCCAACCAAA 1079
                ||||| |||||||| | |
Db    1058 TGAAAGCTTCTTACCCAACTATA 1080
```

RESULT 11
AAC46206
ID AAC46206 standard; DNA; 1538 BP.
XX
AC AAC46206;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49287.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 135.4; DB 21; Length 1538;
Best Local Similarity 50.7%; Pred. No. 1.4e-29;
Matches 480; Conservative 0; Mismatches 431; Indels 36; Gaps 5;

```
Qy    136 TGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAA 195
          |||   ||   |||     |||| || |      |||  ||   | || |||  | |  |
Db    125 TGTATGAAAGGTGGCTTGTTGAGAATCGGAAGAACTATAATGGTCTCGGAGAGAAAGAGA 184

Qy    196 AGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAA 255
          |  ||||||| |||||  | |||| ||| |   ||   ||    | | | | |   |
Db    185 GACGCTTTGAGATCTTCAAAGACAACTTGAAGTTCGTCGAAGAGCACAGTTCCATCCCGA 244

Qy    256 AATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGT 315
          |       |   || |     | || |||| |  ||| |||||  | |||    | ||||
Db    245 A------TCGGACTTACGAAGTCGGGTTGACCCGGTTCGCTGATTTGACTAACGATGAGT 298

Qy    316 TCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACA 375
          |  |    |  ||||||         | |     |  | ||  ||  |   |  |   |
Db    299 TTCGTGCGATTTACTTG---------AGGAGCAAGATGGAGAGAACTAGGGTTCCGGTGA 349

Qy    376 AGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGA 435
          ||  |  |||| |     |  |   | | ||      |||||   |    ||||||||
Db    350 AGGGAGAGAAGTATTTGTATAAGGTTGGAGATAGCTTACCTGATGCTATCGATTGGAGAG 409

Qy    436 AAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTT 495
            |||||||   | |  | ||| ||  | |||||   |||||||||  ||||||||||||
Db    410 CTAAAGGTGCTGTTAATCCAGTCAAAGATCAAGGATCCTGTGGAAGTTGTTGGGCGTTTT 469

Qy    496 CTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTT 555
          | || |  |||||  | ||||    | ||   ||    ||||||||  |  |     | |
Db    470 CCGCGATTGGAGCGGTTGAAGGTATAAATCAGATTAAGACAGGAGAGTTGATATCTTTGT 529

Qy    556 CTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGC--- 612
          |||| |||||||| || || ||||     |     | | | |   |    |||| | |
Db    530 CTGAGCAAGAACTTGTTGATTGTGACACAAGTTACAATGATGGATGTGGTGGAGGTCTTA 589

Qy    613 AGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATC 672
          | || |  | ||  | |   || | ||| | ||||| |||  || || || |||||||
Db    590 TGGATTATGCTTTTAAGTTCATTATTGAAAACGGTGGTATTGATACAGAGGAAGATTATC 649

Qy    673 CTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTG 732
          |||| |  || |  || | ||              | | |   |  |||    |    ||
Db    650 CTTATATTGCAACTGATGTTAACGTGTGCAATTCTGATAAGAAGAACACTCGTGTTGTTA 709

Qy    733 ACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGT 792
          |   |||                 |||||  |        | |   ||| | |  | |
Db    710 CCATCGATG--------------GTTATGAGGATGTTCCTCAAAACGATGAGAAGTCTT 754

Qy    793 TCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTTCATT 852
          |    |  ||  |   | | ||||| ||||||||  | |||||  ||    |       ||
Db    755 TGAAGAAGGCTCTTGCTAATCAACCTATTAGTGTTGCCATTGAAGCTGGTGGCAGAGCTT 814

Qy    853 TATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACT 912
          |  |          |  | || ||     |   | ||| ||          |  | ||
Db    815 TCCAGCTTTAC---ACATCTGGTGTGTTTACAGGAACATGTGGAACATCTTTGGACCATG 871

Qy    913 TTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATT 972
          |||| ||  |||||| || || |||  | ||||  ||| |||||||| |     || |
Db    872 GTGTTGTAGCTGTGGGATACGGATCAGAAGGTGGTCAAGACTACTGGATTGTTCGTAACT 931
```

```
Qy     973 CATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGTAATT 1032
           ||||||||  | | |||||||||    || ||| ||  | || || | ||||      | |
Db     932 CATGGGGATCAAACTGGGGAGAAAGCGGATACTTTAAGCTCGAACGTAACATCAAAGAGT 991

Qy    1033 TATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAA 1079
           || |||    || ||  ||       | ||||| ||||| ||||||
Db     992 CATCAGGGAAATGCGGTGTGGCGATGATGGCTTCTTACCCTACCAAA 1038


RESULT 12
AAF44804
ID   AAF44804 standard; cDNA; 1062 BP.
XX
AC   AAF44804;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Cysteine protease coding sequence #18.
XX
KW   Cell death modulator; programmed cell death; PCD; apoptosis;
KW   forestry plant; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200075331-A1.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-NZ00086.
XX
PR   04-JUN-1999;   99US-0325932.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Flinn B,  Lasham A;
XX
DR   WPI; 2001-061724/07.
DR   P-PSDB; AAB65777.
XX
PT   Novel defender against cell death polynucleotide useful for modulating
PT   programmed cell death pathway and specific development pathways in
PT   forestry plant -
XX
PS   Claim 1; Page 90; 142pp; English.
XX
CC   The present invention relates to coding sequences (see AAF44740-F44840
CC   and AAF44843-F44844) and proteins (see AAB65714-B65814) involved in
CC   programmed cell death (PCD; apoptosis). The coding sequences and proteins
CC   of the present invention are useful for modulating a PCD or cell death
CC   pathway and various developmental pathways in a forestry plant, by
CC   stably incorporating one of the present coding sequences into the genome
CC   of the forestry plant, where the coding sequence provides a PCD pathway
CC   that is not present in a native form of the forestry plant.
XX
SQ   Sequence 1062 BP; 254 A; 270 C; 339 G; 199 T; 0 other;

  Query Match           11.6%;  Score 134.2;  DB 22;  Length 1062;
  Best Local Similarity   51.5%;  Pred. No. 2.7e-29;
  Matches  452;  Conservative    0;  Mismatches  393;  Indels   32;  Gaps    5;

Qy     135 CTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCA 194
           || ||||    |||| ||    ||||       || ||| |  ||  ||  || |  |
Db     194 CTCTTCCGGAGATGGCAGGCGCAGCACAAGAAGGTGTACAAGCACGCCGGGGAGGCCGAG 253

Qy     195 AAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGA 254
           | | | || ||||  ||||| |  |||  ||  ||  | | |||   ||        ||
Db     254 AGGCGGCTCGAGAACTTCAAAAGGAACCTGAGGTACGTGATGGA---GAGAAGCCGGAGG 310

Qy     255 AAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAG 314
           |      | |    ||| |  | ||  ||||||||||| || ||| | |      | |||
Db     311 GATGGCGGCAAGCAGCATGGCGTGGGGCTGAACAAGTTCGCCGACCTGAGCAACGAGGAG 370

Qy     315 TTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAAC 374
           |||||  |    ||| ||          ||| |   |  | | || |     | ||||||
Db     371 TTCAGGCAGCGCTACCTGTCCAAGGTGAAGAAGTCGGTGAACCAGAAGTGGAGGGCCAAG 430

Qy     375 AAGAAAATG-----------AAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCA 422
           | | | |              |||| | |   |       |    | |  | ||  | ||
Db     431 AGGGAGAGCTTGATGAGGAACAAGAGGAAGGGAGCGGAGTCCTGCAAGGCGCCGTCGTCT 490

Qy     423 TGGGATTGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGG 482
           ||||||||||||  | ||  |  |||||   || ||| |||| || |  || |||||
Db     491 CTGGATTGGAGGAACTACGGCATTGTCACCGGCGTGAAGGACCAGGGAGAATGCGGAAGT 550

Qy     483 GGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGAC 542
           | |||||||| ||| | || ||||| || |||| |    ||||  |    |  || |||
Db     551 TGCTGGGCGTTCTCTTCGACCGGAGCAATGGAAGGAATCAATGCGCTCAAGAGCGGGGAC 610

Qy     543 CTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTAC 602
           ||  | ||||||||| || |||||  |||||  |||||  |         | ||| || |   ||
Db     611 CTGATCAGCCTTTCCGAGCAAGAGCTCGTGGACTGCGACACCACCAACGACGGCTGCGAC 670

Qy     603 AATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGAT 662
           ||  |    | |  |  |||||||  |||||  |  |  | || || || |   | ||
Db     671 GGCGGCTACATGGACTATGCGTTCGAGTGGGTCATCAACAACGGCGGTATCGATTCGGAA 730

Qy     663 GATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAG 722
           || || |||||  ||||           ||      |       |   ||  |      |  ||
Db     731 GAAGACTATCCCTACACCAGCGTCTTTGGCGAGGGTGGTATCTGCAACGTCACCAAGGAG 790

Qy     723 GTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACA 782
           | | |         |  |  | | ||    |   |       |  |  | |
Db     791 GAGAACA---------ACAAGGCGGTCACCATTGATGGGTATGTGGACGTTTATCCGTCG 841

Qy     783 GAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGA------T 836
           ||  | |   |  |  || |  || |  | || || ||||| |||   }| ||
Db     842 GACGACGGCCTTCTCTGCACGGTCATCCAACAGCCGATTAGCGTCGGCATGGACGGCTCG 901

Qy     837 GCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCG 896
           || | |||||| ||  | ||||| ||||| || || || ||        |   | | |||
Db     902 GCGATAGATTTCCAGCTCTACACTGGGGGCATCTACGACGGTAGCTGTTCAGCGAATCCC 961

Qy     897 TATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTAC 956
           | |  ||  | ||    ||  |  | || ||||||||  || |  |  || | ||||||
Db     962 GACGACATCGACCATGCGGTCCTGATAGTTGGTTATGGCTCCGAAGGCGGCGAAGATTAT 1021

Qy     957 TGGATAGCGAA--AAATTCATGGGGAGAAGATTGGGG 991
           ||||| | |||  || | | ||||||  ||||| |||
Db    1022 TGGATCGTGAAAGAACTCCTTGGGGAACAGATTTGGG 1058


RESULT 13
AAF44793
ID   AAF44793 standard; cDNA; 1600 BP.
XX
AC   AAF44793;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Cysteine protease coding sequence #7.
XX
KW   Cell death modulator; programmed cell death; PCD; apoptosis;
KW   forestry plant; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200075331-A1.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-NZ00086.
XX
PR   04-JUN-1999;   99US-0325932.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
```

```
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Flinn B,  Lasham A;
XX
DR   WPI; 2001-061724/07.
DR   P-PSDB; AAB65766.
XX
PT   Novel defender against cell death polynucleotide useful for modulating
PT   programmed cell death pathway and specific development pathways in
PT   forestry plant -
XX
PS   Claim 1; Page 86; 142pp; English.
XX
CC   The present invention relates to coding sequences (see AAF44740-F44840
CC   and AAF44843-F44844) and proteins (see AAB65714-B65814) involved in
CC   programmed cell death (PCD; apoptosis). The coding sequences and proteins
CC   of the present invention are useful for modulating a PCD or cell death
CC   pathway and various developmental pathways in a forestry plant, by
CC   stably incorporating one of the present coding sequences into the genome
CC   of the forestry plant, where the coding sequence provides a PCD pathway
CC   that is not present in a native form of the forestry plant.
XX
SQ   Sequence 1600 BP; 390 A; 420 C; 446 G; 344 T; 0 other;

  Query Match            11.5%;  Score 132.8;  DB 22;  Length 1600;
  Best Local Similarity  51.0%;  Pred. No. 8.7e-29;
  Matches  466;  Conservative    0;  Mismatches  412;  Indels   36;  Gaps    5;

Qy     124 AGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACG 183
           |||||    |  | | | |    |||        |||| ||    | |||| |   ||  |
Db     231 AGGTGATGGCCGTCTACGAGAGCTGGCTCGCCAAGCACGGCAAGGCCTACAACGCCCTGG 290

Qy     184 AAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGA 243
             ||  |||  ||| |  |  || | |||||| | |||       |  |||   |||   |
Db     291 GCGAGAAGGAGAAGCGCTTCCAGGTCTTCAAGGACAACCTCCGGTTCATCGACGACCACA 350

Qy     244 ATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCA 303
           | ||   | |  |      ||   |  |     | ||  | ||| |||| || ||| |||
Db     351 ACGCCGGCGGGGA------CCGGACCTACACGGTCGGCCTCAACCAGTTCGCCGACCTCA 404

Qy     304 CTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCA 363
           ||    | |||| | |    | |||| ||   ||    | ||||  ||||  ||
Db     405 CTAACGAGGAGTACCGGTCCATGTACCTGGGCGCCAGGATGGATCGGTCGGGGCGGCGGC 464

Qy     364 AAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCAT 423
             |    |  |       | | |         |  | ||  | |  || || ||
Db     465 TCGGGAGGGCCCGCAGCGATCGGTACGCCGTGGCCGCCGGGGAGGAGCTGCCGGCGTCCG 524

Qy     424 GGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGG 483
           ||||||||||  |||| | | |     | || ||| |||| ||  |||| || ||
Db     525 TCGATTGGAGGAAGGAAGGCGCCGTTGTTGACGTCAAGGACCAGGGAAGCTGCGGGAGTT 584

Qy     484 GTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACC 543
           | |||||||| ||| | |  |  ||  | || |    | |     | |  || || ||
Db     585 GCTGGGCGTTCTCTACAATTGCTGCTGTGGAGGGGATAAACAAGCTTGTGACTGGTGATT 644

Qy     544 TTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAG---AAAGCGAAGGTTCTT 600
           |  |    || || || || ||||| || ||||| |    |    | |  ||||| | |
Db     645 TGATCTCTCTGTCCGAGCAGGAACTTGTGGACTGCGATACATCCTACAATGAAGGATGTA 704

Qy     601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660
           |    ||       | || |  | || ||||   || |  |  | || |||||||  || |
Db     705 ATGGCGGGCTCATGGATTATGCCTTTGAATTCATTATCAACAACGGAGGCATTGATACCG 764

Qy     661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATG---CAAAGCCAATAAGATACAAG 717
           | || |||||||| || ||||||  |||  | |  ||   | ||  ||  ||||    |
Db     765 AGGAAGATTATCCCTATAGAGCTGTAGATAGCACTTGTGACCAATACAGGAAGAACGCAA 824

Qy     718 ACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAG 777
           |    || || |||||||  ||||||  | |   |       |||||||
Db     825 AGGTTGTGACGATTGACGATTATGAAGATGTTCCAGAAAATGATGAGA------------ 872

Qy     778 AGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATG 837
                   |||| ||   ||  ||  |   | | |||||| | |||||  | |||||  |
Db     873 ---------AAGCATTGCAAAAGGCAGTTGCTAATCAACCAGTCAGTGTGGCCATTGAAG 923

Qy     838 CAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGT 897
           ||  ||      | ||  |       |    ||  | |  |   |   | | | ||  |
Db     924 CAGGAGGCCGGGAATTCCAGTTTTATG---ATTCGGGTATATTTACTGGCAAATGTGGGA 980

Qy     898 ATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACT 957
             |   |  ||||    |||      || || |||||  |||   | || || |||||||
Db     981 CAGCTCTGGATCATGGGGTTACTGCAGTCGGATATGGCACAGAAAACGGAGTTGATTACT 1040

Qy     958 GGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAA 1017
           |||||| ||| || |||||||| |     |||||||  | || |||||     ||   |
Db    1041 GGATAGTGAAGAACTCATGGGGCGGTAGCTGGGGAGAGCAAGGGTACATCAAAATGGCAC 1100

Qy    1018 GAAACACGGGTAAT 1031
           ||||   ||  |||
Db    1101 GAAATGTGGCCAAT 1114


RESULT 14
AAC40001
ID   AAC40001 standard; DNA; 1698 BP.
XX
AC   AAC40001;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 26686.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
```

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

```
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             11.3%;  Score 130.8;  DB 21;  Length 1698;
  Best Local Similarity   53.8%;  Pred. No. 3.5e-28;
  Matches  368;  Conservative    1;  Mismatches  293;  Indels   22;  Gaps    4;

Qy     403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
           ||||  | | ||| |  ||    || |||||||| ||||| | | |  |  | || ||
Db     483 GTGATGAGCTACCGGAGTCTATTGACTGGAGGAAGAAAGGAGCCGTGGCTGAGGTCAAAG 542

Qy     463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
           | || || || || || ||  |||||||||||||  |||  |||||  | || | |  |
Db     543 ATCAGGGTGGTTGCGGGAGTTGTTGGGCGTTTTCAACCATTGGAGCAGTGGAGGGAATAA 602

Qy     523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
           |    || | ||| ||||||||  | | | | ||||||||||  | || || ||||
Db     603 ACCAGATCGTAACCGGAGACCTAATAACCTTGTCTGAACAAGAGTTGGTCGATTGTGACA 662

Qy     580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
             | | | |||||||| | ||   ||       | |  |  | || || |   || |
Db     663 CTTCATACAACGAAGGTTGTAACGGAGGTCTTATGGACTATGCTTTTGAGTTCATTATCA 722

Qy     640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
           |  |||||| || |  || ||  | |||||||||||||  | |   || || |  ||
Db     723 AGAATGGTGGAATCGATACAGACAAAGATTATCCTTACAAGGGTGTTGATGGAACTTGTG 782

Qy     700 AAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAG 759
           |    |  | || | | |  || |||   | |  ||   ||  |  |       |  ||
Db     783 ACCAGATCAGGAAAAACGCTAAAGTTGTCACTATCG---ATTCATATGAGGATGTTCCAA 839

Qy     760 ATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAA 819
           | | ||   |||||  || |||       | || |   ||      |     |||
Db     840 CTTACAGCGAGGAATCGTTGAAAGAAAGCAGTTGCTCATCAACCCATTAAGCATCGCCAT 899

Qy     820 TTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAG 879
           | |:         || ||     |  || ||  | ||      ||| || | |||||||
Db     900 TGASG-------CTGGTGGTCGTGCGTTCCAGCTCTATGATTCGGGTATATTTGATGGAA 952

Qy     880 AAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAG 939
                | |||   |         |  | |||   ||| |    || || || ||  | |
Db     953 GTTGTGGAACACAAC---------TAGACCACGGAGTTGTGGCGGTTGGATACGGAACTG 1003

Qy     940 CGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATG 999
           | | ||   ||||||||||||| | || ||| ||||||||  ||   ||||||||  ||
Db    1004 AGAACGGCAAAGATTACTGGATTGTGAGAAACTCATGGGGTAAAAGCTGGGGAGAGAGTG 1063

Qy    1000 GTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATT 1059
           | ||| |  ||||     | ||||  |    || || ||||   ||||| ||      |
Db    1064 GATACCTAAGGATGGCGCGTAACATTGCGTCTTCATCAGGAAAATGTGGAATCGCGATTG 1123

Qy    1060 TCGCTTCATACCCAACCAAAGAGG 1083
              |||||||||| |  ||  | |
Db    1124 AACCTTCATACCCGATAAAGAATG 1147


RESULT 15
AAC51084
ID   AAC51084 standard; DNA; 1369 BP.
XX
AC   AAC51084;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 67216.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX

PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
```

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.1%; Score 128.6; DB 21; Length 1369;
Best Local Similarity 49.6%; Pred. No. 1.4e-27;
Matches 471; Conservative 0; Mismatches 439; Indels 39; Gaps 4;

```
Qy    136 TGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAA 195
          ||||| |   ||||| |    | |||||    || ||| |   |   |  ||  ||| |
Db    244 TGTTCGAGTCATGGATGGTCAAACATGGGAAAGTGTACGACTCCGTTGCCGAGAAGGAAC 303

Qy    196 AGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAA 255
          | |  |   ||||| || | |||      | |||||   ||  ||| ||  | |
Db    304 GGCGTTTAACTATTTTTGAGGACAACCTCCGTTTTATCACTAACCGGAACGCTGAGA--- 360

Qy    256 AATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGT 315
              |    | | ||  | || |||||   |||||| ||  |  |||  || || |
Db    361 ------ATCTCAGTTACCGGCTTGGTTTGAATCGGTTTGCGGATTTATCTCTCCATGAAT 414

Qy    316 TCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACA 375
          | |  |||  | |      |||          |   | |   |   |   |  | |
Db    415 ACGGAGAAATTTGCCATGGGGCTGATCCAAGACCTCCTAGGAACCACGTCTTCATGACTA 474

Qy    376 AGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGA 435
          |  |       |  |       |  | ||||   ||  |||   ||    || ||||| |
Db    475 GCAGCAACCGATACAAGACTAGTGATGGTGATGTTCTTCCTAAGTCCGTTGACTGGAGAA 534

Qy    436 AAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTT 495
          |  |||| |   | ||  |||| ||  | |||||    ||  | ||  |||||||| || |
Db    535 ACGAAGGCGCAGTGACTGAAGTCAAAGATCAAGGCCTTTGCAGGAGTTGTTGGGCTTTCT 594

Qy    496 CTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTT 555
          |  |   ||||||  | ||||    | |    || |  || ||||| || || |   | |
Db    595 CCACTGTGGGAGCAGTGGAAGGCTTAAACAAGATTGTGACTGGAGAGCTAGTAACTTTGT 654
```

```
Qy     556 CTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGT 615
           |||| |||||  |  |  | |||    ||||||| | | ||||        ||
Db     655 CTGAGCAAGATTTGATCAATTGTAACAAAGAAAACAATGGTTGCGGAGGAGGCAAAGTCG 714

Qy     616 ATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTT 675
           |   | | |  || |   |  |  |  |||||||  |||  || ||  | ||||||||||
Db     715 AGACAGCCTATGAGTTCATCATGAACAATGGTGGTCTTGGTACCGACAACGATTATCCTT 774

Qy     676 ACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACG 735
           ||| ||||    | ||    ||| ||| |       |   |||||||
Db     775 ACAAAGCTCTCAATGGAGTCTGCGAAGGCCGCCTCAAGGAAGACAA-------------- 820

Qy     736 GATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCT 795
           |  |   | | ||        ||||||  |        || |    ||    ||  ||
Db     821 -CAAGAATGTTATGATTGATGGGTATGAGAATTTGCCTGCAAACGATGAAGCCGCTCTCA 879

Qy     796 TAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA------AAGATTTTC 849
           | |  ||  |   | | || ||  | | || |    | ||| | |       ||| ||||
Db     880 TGAAAGCGGTTGCTCACCAGCCTGTGACTGCCGTTGTCGATTCCAGCAGCCGAGAGTTTC 939

Qy     850 ATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATC 909
           |  | ||      ||||  | | ||| |||     | | ||||  |         | || |
Db     940 AGCTTTATGAATCGGGAGTGTTTGACGGAACTTGCGGAACAAACC---------TAAACC 990

Qy     910 ACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAA 969
           |   |||| |  | || || ||||  | | | ||||   || ||||||| | |||||
Db     991 ATGGTGTTGTTGTGGTCGGGTATGGAACCGAGAATGGTCGTGACTACTGGATTGTGAAAA 1050

Qy     970 ATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGTA 1029
           | ||  |||| ||    ||||| || | ||| |||||   |||     | ||||  |  |
Db    1051 ACTCGAGGGGCGACACATGGGGGGAGGCTGGCTACATGAAGATGGCTCGCAACATTGCCA 1110

Qy    1030 ATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAA 1078
           ||  |  |||  | ||||| ||          ||||||||||   |||
Db    1111 ATCCAAGAGGCCTATGTGGCATCGCAATGCGAGCTTCATACCCTCTCAA 1159
```

Search completed: January 11, 2003, 19:52:42
Job time : 259 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 19:13:12 ; Search time 2663 Seconds
(without alignments)
12633.437 Million cell updates/sec

Title: US-09-805-694B-1
Perfect score: 1156
Sequence: 1 gcggccgcatgggtttcctt.....tctcctctttgagcggccgc 1156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1156	100.0	1156	6	AX252299	AX252299 Sequence
2	1156	100.0	2970	6	AX252300	AX252300 Sequence
3	1140	98.6	1287	8	SOYP34A	J05560 Glycine max
4	475.4	41.1	5449	8	AB013290	AB013290 Glycine m
5	469.6	40.6	6675	8	AB013289	AB013289 Glycine m
6	251.2	21.7	1732	8	AF182079	AF182079 Matricari
7	167	14.4	1807	6	E13051	E13051 cDNA encodi
8	160.6	13.9	1755	8	STU245924	AJ245924 Solanum t
9	159.4	13.8	1756	8	LEAJ3137	AJ003137 Lycopersi
10	159.4	13.8	1779	8	AF172856	AF172856 Lycopersi
11	155.4	13.4	1751	8	AF259983	AF259983 Ipomoea b
12	155.2	13.4	1699	8	AF454957	AF454957 Brassica
13	154.4	13.4	1056	6	AR049578	AR049578 Sequence
14	154.4	13.4	1056	6	E13053	E13053 Modified cD
15	154.4	13.4	1056	6	E14274	E14274 DNA encodin
16	154.4	13.4	1849	6	E13052	E13052 cDNA encodi
17	151.2	13.1	1577	6	A65336	A65336 Sequence 59
18	151.2	13.1	1577	6	AR150474	AR150474 Sequence
19	150.6	13.0	1680	8	PVZ99954	Z99954 Phaseolus v
20	149.6	12.9	1361	8	VSCYSPRO	X75749 V.sativa L.
21	147	12.7	1553	6	A65337	A65337 Sequence 60
22	147	12.7	1553	6	AR150475	AR150475 Sequence
23	146	12.6	1638	8	AF388175	AF388175 Arabidops
24	144	12.5	1485	8	AY114661	AY114661 Arabidops
25	144	12.5	1628	8	AY062608	AY062608 Arabidops
26	142.2	12.3	1284	8	VSCYSPROA	Z34895 V.sativa L.
27	142	12.3	1389	8	AY081730	AY081730 Arabidops
28	142	12.3	1420	8	AY133781	AY133781 Arabidops
29	142	12.3	1604	8	AY039606	AY039606 Arabidops
30	142	12.3	1656	8	AY072130	AY072130 Arabidops
31	141.4	12.2	1351	8	PVZ99952	Z99952 Phaseolus v
32	141.4	12.2	1382	8	PVA224766	AJ224766 Phaseolus
33	141.4	12.2	1382	8	PVU52970	U52970 Phaseolus v
34	139.2	12.0	1219	8	AF191028	AF191028 Arabidops
35	138	11.9	1041	8	AY040073	AY040073 Arabidops
36	137	11.9	1218	8	NTCYSTPRO	Z99173 Nicotiana t
37	136.6	11.8	1253	8	AF370131	AF370131 Arabidops
38	135.4	11.7	1414	8	AY093350	AY093350 Arabidops
39	135.4	11.7	1575	8	AY062725	AY062725 Arabidops
40	135	11.7	1425	8	AY043294	AY043294 Arabidops
41	133.6	11.6	1314	8	VMSHEP	X15732 Vigna mungo
42	133.2	11.5	1830	8	AF454958	AF454958 Brassica
43	132	11.4	1221	8	PVEPC1	X56753 Phaseolus v
44	131.8	11.4	1820	8	PSTHIOL	X66061 P.sativum m
45	130.8	11.3	1221	8	VRU49445	U49445 Vigna radia

ALIGNMENTS

RESULT 1
AX252299
LOCUS AX252299 1156 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 1 from Patent WO0168887.
ACCESSION AX252299
VERSION AX252299.1 GI:15985640
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 1156)
AUTHORS Jung,R. and Kinney,A.J.
TITLE Hypoallergenic transgenic soybeans
JOURNAL Patent: WO 0168887-A 1 20-SEP-2001;
E. I. du Pont de Nemours and Company (US) ; PIONEER HI-BRED

```
           INTERNATIONAL, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chimeric construct"
BASE COUNT      361 a    223 c    277 g    295 t
ORIGIN

  Query Match             100.0%;  Score 1156;  DB 6;  Length 1156;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-290;
  Matches 1156;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 60

Qy     61 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 120

Qy    121 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 180

Qy    181 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 240

Qy    241 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 300

Qy    301 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 360

Qy    361 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 420

Qy    421 CATGGGATTGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 480

Qy    481 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 540

Qy    541 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 600

Qy    601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660

Qy    661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 720

Qy    721 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 780

Qy    781 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 840

Qy    841 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 900

Qy    901 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 960

Qy    961 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1020

Qy   1021 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1080

Qy   1081 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1140

Qy   1141 CTCTTTGAGCGGCCGC 1156
          ||||||||||||||||
Db   1141 CTCTTTGAGCGGCCGC 1156


RESULT 2
AX252300
LOCUS       AX252300               2970 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 2 from Patent WO0168887.
ACCESSION   AX252300
VERSION     AX252300.1  GI:15985641
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 2970)
  AUTHORS   Jung,R. and Kinney,A.J.
  TITLE     Hypoallergenic transgenic soybeans
  JOURNAL   Patent: WO 0168887-A 2 20-SEP-2001;
            E. I. du Pont de Nemours and Company (US) ; PIONEER HI-BRED
            INTERNATIONAL, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .2970
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chimeric construct"
BASE COUNT     1018 a    547 c    527 g    878 t
ORIGIN

  Query Match             100.0%;  Score 1156;  DB 6;  Length 2970;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-290;
  Matches 1156;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    622 GCGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTT 681

Qy     61 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    682 CCAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGA 741

Qy    121 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    742 AACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACC 801

Qy    181 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    802 ACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACA 861

Qy    241 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    862 TGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACA 921

Qy    301 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    922 TCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAA 981
```

```
Qy     361 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     982 TCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCAT 1041

Qy     421 CATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1042 CATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAA 1101

Qy     481 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1102 GGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAG 1161

Qy     541 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1162 ACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTT 1221

Qy     601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1222 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 1281

Qy     661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1282 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 1341

Qy     721 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1342 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 1401

Qy     781 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1402 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 1461

Qy     841 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1462 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 1521

Qy     901 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1522 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 1581

Qy     961 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1582 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1641

Qy    1021 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1642 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1701

Qy    1081 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1702 AGGAATCAGAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTC 1761

Qy    1141 CTCTTTGAGCGGCCGC 1156
           ||||||||||||||||
Db    1762 CTCTTTGAGCGGCCGC 1777
```

```
RESULT 3
SOYP34A
LOCUS       SOYP34A                 1287 bp    mRNA    linear   PLN 30-SEP-1996
DEFINITION  Glycine max 34 kDa maturing seed vacuolar thiol protease mRNA,
            complete cds.
ACCESSION   J05560
VERSION     J05560.1  GI:1199562
KEYWORDS    soybean protein storage vacuole thiol protease.
SOURCE      Glycine max.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Kalinski,A., Weisemann,J.M., Matthews,B.F. and Herman,E.M.
  TITLE     Molecular cloning of a protein associated with soybean seed oil
            bodies that is similar to thiol proteases of the papain family
  JOURNAL   J. Biol. Chem. 265 (23), 13843-13848 (1990)
  MEDLINE   90338001
   PUBMED   2380191
REFERENCE   2  (bases 1 to 1287)
  AUTHORS   Kalinski,A., Melroy,D.L., Dwivedi,R.S. and Herman,E.M.
  TITLE     A soybean vacuolar protein (P34) related to thiol proteases is
            synthesized as a glycoprotein precursor during seed maturation
  JOURNAL   J. Biol. Chem. 267 (17), 12068-12076 (1992)
  MEDLINE   92291086
   PUBMED   1601876
REFERENCE   3  (bases 1 to 1287)
  AUTHORS   Herman,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-1990) Eliot M. Herman, Plant Molecular Biology
            Laboratory, USDA, Agricultural Research Service, Beltsville, MD
            20705
COMMENT     On Feb 21, 1996 this sequence version replaced gi:170039.
FEATURES             Location/Qualifiers
     source          1. .1287
                     /organism="Glycine max"
                     /strain="Century"
                     /db_xref="taxon:3847"
                     /clone="P34/1350"
                     /tissue_type="seed"
     mRNA            <1. .1287
     CDS             3. .1142
                     /codon_start=1
                     /product="34 kDa maturing seed vacuolar thiol protease
                     precursor"
                     /protein_id="AAB09252.1"
                     /db_xref="GI:1199563"
                     /translation="MGFLVLLLFSLLGLSSSSSISTHRSILDLDLTKFTTQKQVSSLF
                     QLWKSEHGRVYHNHEEEAKRLEIFKNNSNYIRDMNANRKSPHSHRLGLNKFADITPQE
                     FSKKYLQAPKDVSQQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQVKYQGGCGRGW
                     AFSATGAIEAAHAIATGDLVSLSEQELVDCVEESEGSYNGWQYQSFEWVLEHGGIATD
                     DDYPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSID
                     AKDFHLYTGGIYDGENCTSPYGINHFVLLVGYGSADGVDYWIAKNSWGEDWGEDGYIW
                     IQRNTGNLLGVCGMNYFASYPTKEESETLVSARVKGHRRVDHSPL"
     sig_peptide     3. .368
     mat_peptide     369. .1139
                     /product="34 kDa maturing seed vacuolar thiol protease"
BASE COUNT      407 a    241 c    297 g    342 t
ORIGIN
```

```
  Query Match             98.6%;  Score 1140;  DB 8;  Length 1287;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-286;
  Matches 1140;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 68
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 62

Qy      69 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 122

Qy     129 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 182

Qy     189 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 242

Qy     249 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 302

Qy     309 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 362
```

```
Qy     369 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 422

Qy     429 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGG 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGG 482

Qy     489 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 542

Qy     549 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA 608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA 602

Qy     609 TGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 TGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT 662

Qy     669 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA 728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA 722

Qy     729 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA 788
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     723 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA 782

Qy     789 GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTT 848
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     783 GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTT 842

Qy     849 CATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAAT 908
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     843 CATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAAT 902

Qy     909 CACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAA 968
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     903 CACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAA 962

Qy     969 AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT 1028
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     963 AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT 1022

Qy    1029 AATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAGAGGAATCA 1088
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1023 AATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAGAGGAATCA 1082

Qy    1089 GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTTGA 1148
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1083 GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTTGA 1142


RESULT 4
AB013290
LOCUS       AB013290               5449 bp    DNA     linear   PLN 29-APR-1998
DEFINITION  Glycine max pseudogene for Bd 30K.
ACCESSION   AB013290
VERSION     AB013290.1  GI:3097322
KEYWORDS    pseudogene; Bd 30K.
SOURCE      Glycine max DNA.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 5449)
  AUTHORS   Takano,T. and Yamada,M.
  TITLE     Cloning of a pseudo gene for an allergenic protein Gly m Bd 30K in
            soybean
  JOURNAL   Published Only in DataBase (1998)
REFERENCE   2  (bases 1 to 5449)
  AUTHORS   Takano,T. and Yamada,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1998) Tetsuo Takano, University of Tokyo, Asian
            Natural Environmental Science Center; University Farm, Univ. Tokyo
            1-1-1 Midori-cho, Tanashi, Tokyo 188-0002, Japan
            (E-mail:takano@ims.u-tokyo.ac.jp, Tel:+81-424-63-1618,
            Fax:+81-424-63-1618)
FEATURES             Location/Qualifiers
     source          1. .5449
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
     misc_feature    1. .5449
                     /note="pseudogene for an allergenic protein Gly m Bd 30K"
                     /pseudo
BASE COUNT     1823 a    917 c    862 g   1847 t
ORIGIN

  Query Match             41.1%;  Score 475.4;  DB 8;  Length 5449;
  Best Local Similarity   71.7%;  Pred. No. 6.4e-113;
  Matches  802;  Conservative    0;  Mismatches  101;  Indels  216;  Gaps    5;

Qy       9 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 68
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3421 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 3480

Qy      69 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 128
           |   |||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db    3481 T---CTCATCGTTCCATATTGGACCTTGACCTAGCCAAGTTCACCACACAGAAACAGGTG 3537

Qy     129 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 188
           |||||||||||||||| |||||||||| |||| ||||| |||||||||||||| ||||||
Db    3538 TCTTCACTGTTCCAACCATGGAAGAGTAAGCACGGACGGGTCTACCATAACCAAGAAGAA 3597

Qy     189 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 248
           ||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||
Db    3598 TAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTGTATCAGGGACATGAATGCA 3657

Qy     249 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 308
           |||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||
Db    3658 AACAGAAAATCACCCCACTCTCATCGTTTGGGATTGAACAAGTTTGCTGACATCACTCCT 3717

Qy     309 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 368
           |||||||||||||||||| |||||||||||||||||||| | |  || ||||| |||
Db    3718 CAAGAGTTCAGCAAAAAGTATTTGCAAGCTCCCAAGGATGTGCCACGACACATCAACATG 3777

Qy     369 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 428
           ||| ||||| || |||||  ||||||||||||||||||||||||||||||||||||||||
Db    3778 GCCGACAAGGAATTGAAGGAGGAACAGCATTCTTGTGACCATCCACCTGCATCCTGGGAT 3837

Qy     429 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGT------------ 476
           |||||| |||| |||||||||||  | |||||| |||||||    |||
Db    3838 TGGAGGGAAAAGGGTGTCATCACGGATGTAAAGCACCAAGGCTTGTGTGGTACGTGCATG 3897

Qy     477 ------------------------------------------------------------ 476

Db    3898 AAACCATTGTTTTTAGTACGTAATTACTAATTATTAACACTAGTATTTTTTTCTCTCTTT 3957

Qy     477 ----------------------------------GGAAGGGGTTGGGCGTTTTCTGCCA 501
                                               || || || ||||| ||| ||||||
Db    3958 GAAAATGACTTGTGGAGTGTGAAACCATTTTGATAGGGAGTGGGTGGGCATTTGCTGCCA 4017

Qy     502 CGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAAC 561
           |||||||||| ||||| | |||||||||||||||||||||||||| ||||||||||||||
Db    4018 CGGGAGCCATTGAAGCGGTACATGCAATAGCAACAGGAGACCTTG-TAGCCTTTCTGAAC 4076

Qy     562 AAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAAT 621
           |||||||||||| ||||||   |||||||||||| || || ||||||||   || | |
Db    4077 AAGAACTCGTAGATTGTGTG---GAAAGCGAAGGTTGTTCCAGTGGATGGCCCGATGATT 4133

Qy     622 CGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAG 681
           |||||| |||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db    4134 CGTTCGTATGGGTTTTAGAACATGGTGGGATTGCCACTGATGCTGATTATCCTTACAGAG 4193
```

```
Qy     682 CTAAAGAGGGTAGATGCAAAGCCAATA------------------------------ 708
           |||||||| ||||||||||||||||||
Db    4194 CTAAAGAGTGTAGATGCAAAGCCAATAAGGTCTGTGAATATATAGCTGAATATGCATGCA 4253

Qy     709 ------------------------------------------------------------ 708

Db    4254 TGCAATTCCTTTTGATTAATTTAACCCTTCATCTCTTTGTTTTGTGCACTTTTTTTTCTT 4313

Qy     709 ---------AGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAG 759
                   |||||||||||||||||||||||||||||||| | | ||| |||| ||
Db    4314 TCTGGGAGTAGATACAAGACAAGGTTACAATTGACGGATATGACATTATAAGAATGCCAT 4373

Qy     760 ATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAA 819
           |||||||||||||||||||| |||| |||||||||| ||||||||||||||||||||||
Db    4374 ATGAGAGTACAGAATCAGAGAGAGAGAAAGCGTTCTTATGCGCCATCCTTGAGCAACCAA 4433

Qy     820 TTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAG 879
           |||||||||| |||||||| |||||||||| | ||||| |||| || | |  |    |
Db    4434 TTAGTGTCTCCATTGATGCAGAAGATTTTCATCTCTACACTGGGGTAAATAAGTAATATG 4493

Qy     880 AAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTT 918
             |  ||||   |     || | | |   | || || | |
Db    4494 TACTCTGTTTCAAAAAATAAGTGTTATCTTACATTTTAT 4532
```

RESULT 5
AB013289
LOCUS AB013289 6675 bp DNA linear PLN 29-APR-1998
DEFINITION Glycine max gene for Bd 30K, complete cds.
ACCESSION AB013289
VERSION AB013289.1 GI:3097320
KEYWORDS Bd 30K.
SOURCE Glycine max DNA.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 6675)
AUTHORS Takano,T. and Yamada,M.
TITLE Cloning of a gene for an allergenic protein in soybean
JOURNAL Published Only in DataBase (1998)
REFERENCE 2 (bases 1 to 6675)
AUTHORS Takano,T. and Yamada,M.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1998) Tetsuo Takano, University of Tokyo, Asian Natural Environmental Science Center; University Farm, Univ. Tokyo 1-1-1 Midori-cho, Tanashi, Tokyo 188-0002, Japan (E-mail:takano@ims.u-tokyo.ac.jp, Tel:+81-424-63-1618, Fax:+81-424-63-1618)
FEATURES Location/Qualifiers
source 1. .6675
/organism="Glycine max"
/db_xref="taxon:3847"
CDS join(4647. .5115,5474. .5706,5820. .5972,6178. .6462)
/codon_start=1
/product="Bd 30K"
/protein_id="BAA25899.1"
/db_xref="GI:3097321"
/translation="MGFLVLLLFSLLGLSSSSSISTHRSILDLDLTKFTTQKQVSSLF
QLWKSEHGRVYHNHEEEAKRLEIFKNNLNYIRDMNANRKSPHSHRLGLNKFADITPQE
FSKKYLQAPKDVSQQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQVKYQGGCGSGW
AFSATGAIEAAHAIATGDLVSLSEQELVDCVEESEGCYNGWHYQSFEWVLEHGGIATD
DDYPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSID
AKDFHLYTGGIYDGENCTSPYGINHFVLLVGYGSADGVDYWIAKNSWGEDWGEDGYIW
IQRNTGNLLGVCGMNYFASYPTKEESETLVSARVKGHRRVDHSPL"
BASE COUNT 2349 a 1024 c 1074 g 2228 t
ORIGIN

Query Match 40.6%; Score 469.6; DB 8; Length 6675;
Best Local Similarity 99.2%; Pred. No. 2e-111;
Matches 472; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       9 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 68
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4647 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 4706

Qy      69 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4707 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 4766

Qy     129 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4767 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 4826

Qy     189 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 248
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    4827 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCA 4886

Qy     249 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4887 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 4946

Qy     309 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4947 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 5006

Qy     369 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5007 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 5066

Qy     429 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGG 484
           ||||||||||||||||||||||||||||||||||||||||||||||||||| | | |
Db    5067 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGTATGTG 5122
```

RESULT 6
AF182079
LOCUS AF182079 1732 bp mRNA linear PLN 10-SEP-1999
DEFINITION Matricaria chamomilla thiol protease (ctp) mRNA, complete cds.
ACCESSION AF182079
VERSION AF182079.1 GI:5853328
KEYWORDS .
SOURCE Matricaria chamomilla.
ORGANISM Matricaria chamomilla
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae; Anthemideae; Matricaria.
REFERENCE 1 (bases 1 to 1732)
AUTHORS Kohchi,C., Yasuda,M. and Hirata,T.
TITLE Isolation of cDNA encoding thiol protease from cultured shoot primordia of Matricaria chamomilla
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1732)
AUTHORS Kohchi,C., Yasuda,M. and Hirata,T.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526, Japan
FEATURES Location/Qualifiers
source 1. .1732
/organism="Matricaria chamomilla"
/db_xref="taxon:98504"
/tissue_type="cultured shoot primordia"
gene 1. .1732
/gene="ctp"
CDS 44. .1549
/gene="ctp"
/codon_start=1
/product="thiol protease"
/protein_id="AAD54424.1"
/db_xref="GI:5853329"
/translation="MATSNSMITILIFLTYVSYSISTKTLPSEFSILEGQENDILSSA
KVSDLFGKWKELHGKTYQHEEEENLRLENFKKSVKFVMEKNSERKSELDHTVGLNKFA
DLSNEEFKEMYMSKVKGSRSNELKMGGVKRNMSVSSRTCDAPTSLDWRDKGVVTPMKD

```
                    QGQCGSCWAFSVSGSIESANAIATGDLIRLSEQELVDCDTYDYGCDGGNMDTAYRWII
                    KNGGLDSEDDYPYTSSNGRDGKCDKTKSAKSVVSLDSYVEVESNEDAVLCAVATTPVT
                    IGIVGSAYDFQLYTGGVYNGQCSSKPYDIDHAVLIVGYGSQDGKDYWIVKNSWGTYWG
                    LEGYILMERNTDIKNGVCGMYLEPVYPITAAPTPPGPPPPPAPPSPPHPPPPPTPPAP
                    SKCGDFHYCAADQTCCCIFEFYNYCLIYGCCGYSDAVCCKNSAACCPSDYPICDVQAG
                    YCYKNSAKTFGVPAKKRQLAKHKMPWEKIEETIKEEFQPLAWNRNPFAAAA"
BASE COUNT      553 a    319 c    403 g    457 t
ORIGIN

  Query Match             21.7%;  Score 251.2;  DB 8;  Length 1732;
  Best Local Similarity   55.4%;  Pred. No. 1.6e-54;
  Matches  572;  Conservative    0;  Mismatches  448;  Indels   12;  Gaps    4;

Qy      70 CAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGT 129
           | | | |   ||| ||| | ||   | |   ||   |  ||    ||     | | ||
Db     123 CTAGTGAGTTTTCTATACTTGAAGGTCAAGAAAATGATATTCTATCAAGTGCAAAAGTTA 182

Qy     130 CTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAG 189
           |   || ||   |  ||||||     |||||||     |||||  |  | |||||  |
Db     183 GTGACCTATTTGGAAAATGGAAAGAGTTGCATGGAAAAACATACCAACATGAAGAAGAGG 242

Qy     190 AGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAA 249
           ||      || ||||||| |||||||| |     || | | | | ||| |  ||  ||
Db     243 AGAACCTTAGGCTTGAGAATTTCAAGAAAAGTGTAAAGTTTGTAATGGAGAAAAACTCAG 302

Qy     250 ACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTC 309
           | || || ||    |    ||||  | | |||||||| ||||||||||| | | |  |
Db     303 AGAGGAAGTCTGAGCTAGATCATACTGTGGGATTGAATAAGTTTGCTGACTTGAGTAATG 362

Qy     310 AAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGG---ATGTGTCGCAGCAAATCAAAA 366
           |||||||  |   | | |||  ||  |    ||||    |          |  || | || |
Db     363 AAGAGTTTAAAGAGATGTATATGTCAAAAGTTAAGGGGTCTAGAAGTAATGAATTAAAGA 422

Qy     367 TGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTG---TGACCATCCACCTGCATCAT 423
           |||       ||||| ||| | |     |  |||  |   |    || ||||  | || |
Db     423 TGGGGGGTGTGAAAAGGAACATGAGTGTAAGTTCGAGGACTTGTGATGCACCAACTTCTT 482

Qy     424 GGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGG 483
           |||||||||  | ||||| ||  | || | | | ||| |||||||     ||||||||
Db     483 TGGATTGGAGAGACAAAGGAGTAGTTACACCAATGAAGGACCAAGGCCAGTGTGGAAGTT 542

Qy     484 GTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACC 543
           ||||||| || || |   | ||| | || |||   ||| ||||||||| || || ||||
Db     543 GTTGGGCATTCTCAGTGTCTGGATCAATCGAAAGTGCAAATGCAATAGCCACCGGTGACC 602

Qy     544 TTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACA 603
           |  |||| || || || ||||| || || |||||||      |   | | || |   |
Db     603 TCATTAGACTCTCAGAGCAAGAGCTTGTCGACTGTGACACTTATGACTATGGGTGCGATG 662

Qy     604 ATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATG 663
           ||||     | |     | |    |||| || |  |  |||||||  |||   |||| |
Db     663 GTGGAAACATGGACACGGCTTATAGATGGATTATTAAGAATGGTGGGCTTGATTCTGAAG 722

Qy     664 ATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGG 723
           | |||||||| ||||    |    | || |      |    | | |  |||| |||| |
Db     723 ACGATTATCCATACACCAGTTCTAATGGCCG----CGATGGTAAATGTGACAAAACAAAG 778

Qy     724 TTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAG 783
           | |  | | | | |    || |    | |  |     || |  |||| |
Db     779 TCAGCAAAGTCAGTTGTCTCGCTTGATAGCTATGTGGAAGTAGAATCGAATGAAGATGCA 838

Qy     784 AGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAG 843
             || ||   ||  |  ||      |  | ||| || ||     |   ||  | |
Db     839 GTTTTATGTGCTGTGGCAACTACCCCTGTTA--CTATTGGTATCGTGGGTTCGGCTTATG 896

Qy     844 ATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGA 903
           |||| ||  |||||||||||||  | ||| | ||| ||  ||  |  |  || |||| |
Db     897 ATTTCCAGCTATACACCGGGGGTGTATATAACGGACAATGCTCAAGTAAACCATATGATA 956

Qy     904 TTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAG 963
           |  | ||    ||  || | || ||||||||||||   |||||    || ||||||| |
Db     957 TAGACCATGCAGTGCTAATCGTTGGTTATGGTTCACAAGATGGCAAGGACTACTGGATCG 1016

Qy     964 CGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACA 1023
             || |||||||||||     ||||||||  ||| || ||||||   ||  |||||||||
Db    1017 TCAAGAATTCATGGGGCACTTATTGGGGACTAGAAGGCTACATTCTAATGGAAAGAAACA 1076

Qy    1024 CGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAGAGG 1083
           | | || |  |   || || ||||| ||| |   |    |    |||||  | ||  | ||
Db    1077 CTGATATTAAAAACGGGGTTTGTGGAATGTACCTTGAGCCGGTGTACCCTATCACTGCGG 1136

Qy    1084 AATCAGAAACAC 1095
            ||  | | |
Db    1137 CGCCAACACCCC 1148


RESULT 7
E13051
LOCUS       E13051                  1807 bp    DNA     linear   PAT 27-APR-1998
DEFINITION  cDNA encoding thiol protease D3-alpha.
ACCESSION   E13051
VERSION     E13051.1  GI:3251863
KEYWORDS    JP 1997121870-A/1.
SOURCE      Glycine max.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 1807)
  AUTHORS   Kawai,M., Asano,M., Suzuki,S., Miwa,T., Shibai,H. and Yokozeki,K.
  TITLE     DNA CAPABLE OF CODING NEW THIOL PROTEASE AND PRODUCTION OF THE SAME
            THIOL PROTEASE USING THE DNA
  JOURNAL   Patent: JP 1997121870-A 1 13-MAY-1997;
            AJINOMOTO CO INC
COMMENT     OS   Glycine max
            PN   JP 1997121870-A/1
            PD   13-MAY-1997
            PF   28-DEC-1995 JP 1995353931
            PR   29-DEC-1994 JP 94P   340399, 30-AUG-1995 JP 95P   245279 PI
            KAWAI MISAKO, ASANO MINAO, SUZUKI SHUNICHI, MIWA TETSUYA, PI
            SHIBAI HIROSHIRO, YOKOZEKI KENZO
            PC   C12N15/09,C12N1/19,C12N1/21,C12N9/50,(C12N1/19,C12R1:865), PC
             (C12N1/21,
            PC   C12R1:19),(C12N9/50,C12R1:19),(C12N9/50,C12R1:865); CC
            strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1807
            FT                   /organism='Glycine max'
            FT   5'UTR           1. .67
            FT   sig_peptide     68. .460
            FT                   /product='signal peptide of thiol protease  FT
                                D3-alpha'
            FT   mat_peptide     461. .1456
            FT                   /product='thiol protease D3-alpha' FT   CDS
                 68. .1459
            FT                   /product='thiol protease D3-alpha' FT   3'UTR
                 1460. .1807.
FEATURES            Location/Qualifiers
     source         1. .1807
                    /organism="Glycine max"
                    /db_xref="taxon:3847"
BASE COUNT      500 a    387 c    449 g    471 t
ORIGIN

  Query Match             14.4%;  Score 167;  DB 6;  Length 1807;
  Best Local Similarity   53.0%;  Pred. No. 1.3e-32;
  Matches  493;  Conservative    0;  Mismatches  405;  Indels   33;  Gaps    5;

Qy     103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
           |||  ||    ||   |  |  || ||   ||  ||| | | |  |||  |    |||| |
Db     168 CCACGTCGCGCAGCGACGAGGAGCTGATGTCCATGTACGAGCAGTGGCTGGTGAAGCACG 227
```

```
Qy     163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
           |    || ||| |    | ||  ||  |||  |||||  |  |||| |||||| | |||
Db     228 GCAAGGTGTACAACGCGCTCGGGGAGAAGGAGAAGAGGTTCCAGATCTTCAAGGACAACC 287

Qy     223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
           |   | |||   |||   ||  | |     |   | || |   |     | |||
Db     288 TGCGATTCATCGACGACCACAACTCCCAAGAGGACCGAACCTACAAGC------TCGGAC 341

Qy     283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
           ||||| ||| || ||  ||||     | |||| |||    |||||||||  | |    |
Db     342 TGAACCGGTTCGCGGATCTCACCAACGAGGAGTACAGGGCCAAGTACTTGGGAACCAAGA 401

Qy     343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
           |||      | |    ||   | | ||   |  ||     |   |  ||    |
Db     402 TCGATCCCAACCGGAGGCTCGGCAAGACCCCGAGCAACCGATACGCGCCACG---TGTCG 458

Qy     403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
           | ||| | | ||||| |||    |||||||||| ||||||   |  | | ||| ||
Db     459 GCGACAAACTACCTGAATCGGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCAGTCAAAG 518

Qy     463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
           ||||||| |||||||| ||  |||||||  || || || |  || ||  | |||| |  |
Db     519 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTGGAAGGAATAA 578

Qy     523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
           ||   ||||  ||||| || ||  ||     | || |||||||||  | ||  || ||||
Db     579 ATAAGATAGTGACAGGTGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 638

Qy     580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
           | | | | |||||| |   |    |||     | |   |  | || || |   || |
Db     639 CAGGATATAACGAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATTATCA 698

Qy     640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
           |  |||| |||||||   |||| || ||||| |||||| | | |   || |||||||||
Db     699 ACAATGGCGGCATTGATTCTGAAGAGGATTACCCTTACCGTGGTGTTGATGGTAGATGTG 758

Qy     700 AAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAG 759
           |  |  ||| || | | |  ||||||     |   |   |  |         |     |
Db     759 ACACATATAGGAAAAATGCTAAGGTTGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 818

Qy     760 ATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAA 819
           ||||                  |||  ||| ||   ||  ||  |     | || ||
Db     819 ATGAT-----------------GAGTTAGCTTTGAAAAAGGCTGTTGCAAATCAGCCCG 860

Qy     820 TTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAG 879
           | ||||| | ||||| | |    |       | ||  |        |   | | || || |
Db     861 TCAGTGTAGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTGTAT 917

Qy     880 AAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAG 939
           |   | | | ||      |   |  ||||   |||  |   ||| || ||||||  |||
Db     918 TCACTGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGTACAG 977

Qy     940 CGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATG 999
           |  |||||   ||||| ||||| | |  ||||||||||||       ||||||||||||
Db     978 CTAATGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTCCTAGCTGGGGAGAAGATG 1037

Qy    1000 GTTACATTTGGATCCAAAGAAACACGGGTAA 1030
           | |||||  || |  |||||||    | |||
Db    1038 GCTACATCAGGTTAGAAAGAAATCTTGCTAA 1068
```

RESULT 8
STU245924

```
LOCUS       STU245924              1755 bp    mRNA    linear   PLN 16-FEB-2000
DEFINITION  Solanum tuberosum mRNA for cysteine protease (cyp gene).
ACCESSION   AJ245924
VERSION     AJ245924.1  GI:5777888
KEYWORDS    CyP gene; cysteine protease.
SOURCE      Solanum tuberosum.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 1755)
  AUTHORS   Avrova,A.O., Stewart,H.E., De Jong,W.D., Heilbronn,J., Lyon,G.D.
            and Birch,P.R.
  TITLE     A cysteine protease gene is expressed early in resistant potato
            interactions with Phytophthora infestans
  JOURNAL   Mol. Plant Microbe Interact. 12 (12), 1114-1119 (1999)
  MEDLINE   20089145
   PUBMED   10624019
REFERENCE   2  (bases 1 to 1755)
  AUTHORS   Avrova,A.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Avrova A.O., Fungal and Bacterial Plant
            Pathology, Scottish Crop Research Institute, Invergowrie, DD2 5DA,
            UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .1755
                     /organism="Solanum tuberosum"
                     /cultivar="Stirling"
                     /db_xref="taxon:4113"
                     /tissue_type="leaf"
                     /country="United Kingdom"
     gene            1. .1401
                     /gene="cyp"
     CDS             1. .1401
                     /gene="cyp"
                     /function="proteolysis"
                     /codon_start=1
                     /product="cysteine protease"
                     /protein_id="CAB53515.1"
                     /db_xref="GI:5777889"
                     /translation="MAAHSSTLTISLLLMLIFSTLSSASDMSIISYDETHIHHRSDDE
                     VSALYESWLIEHGKSYNALGEKDKRFQIFKDNLKYIDEQNSVPNQSYKLGLTKFADLT
                     NEEYRSIYLGTKSSGDRRKLSKNKSDRYLPKVGDSLPESVDWRDKGVLVGVKDQGSCG
                     SCWAFSAVAAMESINAIVTGNLISLSEQELVDCDKSYNEGCDGGLMDYAFEFVINNGG
                     IDTEEDYPYKERNDVCDQYRKNAKVVKIDSYEDVPVNNEKALQKAVAHQPVSIAIEAG
                     GRDLQHYKSGIFTGKCGTAVDHGVVAAGYGSENGMDYWIVRNSWGAKWGEKGYLRVQR
                     NVASSSGLCGLATEPSYPVKTGANPPKPAPSPPSPVKPPTECDEYSQCPVGTTCCVL
                     EFRRSCFSWGCCPLEGATCCEDHSSCCPHDYPVCNVRQGTCSMSKGNPLGVKAMKRIL
                     AQPIGAFGNGGKKSSS"
BASE COUNT      518 a    350 c    411 g    476 t
ORIGIN
```

```
  Query Match             13.9%;  Score 160.6;  DB 8;  Length 1755;
  Best Local Similarity   51.0%;  Pred. No. 6e-31;
  Matches  503;  Conservative    0;  Mismatches  454;  Indels   30;  Gaps    4;

Qy     124 AGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACG 183
           | || ||  |  ||| | |    |||   | |||||| |||      ||| |       |
Db     131 AAGTTTCGGCGTTGTACGAGTCCTGGCTAATTGAGCACGGAAAATCGTACAACGCTTTAG 190

Qy     184 AAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGA 243
           |||| |||  |||||| |  |||| |||||| |||||| ||| |||||    ||   ||
Db     191 GAGAAAAGGATAAGAGATTCCAGATCTTCAAGGATAACTTGAAATATATAGACGAACAGA 250

Qy     244 ATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCA 303
           | |       ||       ||   | ||    | |||||||  || || || ||  | |
Db     251 ACTCTGTTCCGAAT------CAGAGTTATAAGCTCGGATTGACGAAATTCGCCGATCTGA 304

Qy     304 CTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCA 363
           ||  | | |||| |||    |  ||||||    ||   |     |     | |   |
Db     305 CTAATGAGGAGTACAGGTCGATTTACTTGGGCACTAAGAGTTCCGGTGACCGGAGGAAGT 364

Qy     364 AAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCAT 423
           |   ||||| |       | |    |   |  | | ||       || | ||||
Db     365 TGTCGAAGAACAAAAGTGATCGGTATCTTCCTAAAGTTGGGGATAGCTTGCCGGAATCAG 424

Qy     424 GGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGG 483
           || |||||  | ||||||||  |       || ||| | |||||  |||||| ||
Db     425 TTGACTGGAGAGATAAAGGTGTGCTTGTTGGTGTGAAGGATCAAGGAAGCTGTGGGAGTT 484

Qy     484 GTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACC 543
           ||||||| || ||||||   | |||||| ||| ||  | |||| |||| ||| ||  |
```

```
Db    485 GTTGGGCATTCTCTGCCGTTGCTGCCATGGAATCAATAAATGCGATAGTAACCGGGAATT 544

Qy    544 TTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGCGAAGGTTCTT 600

Db    545 TGATATCACTATCAGAGCAAGAGTTGGTGGATTGTGATAAGTCGTACAATGAAGGTTGCG 604

Qy    601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660

Db    605 ATGGTGGTCTTATGGACTACGCCTTTGAATTCGTCATCAATAATGGAGGAATTGACACTG 664

Qy    661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACA 720

Db    665 AAGAAGACTACCCTTACAAAGAACGTAATGACGTATGCGACCAATATAGGAAAAATGCCA 724

Qy    721 AGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGA 780

Db    725 AGGTTGTTAA----------------AATAGATAGCTATGAAGATGTTCCTGTTAATA 766

Qy    781 CAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA 840

Db    767 ACGAAAAGGCGCTACAAAAGGCTGTTGCACATCAACCTGTGAGCATTGCAATTGAAGCTG 826

Qy    841 AAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATG 900

Db    827 GTGGCAGAGACTTACAGCAC---TACAAATCTGGTATCTTTACTGGAAAATGTGGTACTG 883

Qy    901 GGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGA 960

Db    884 CAGTGGATCATGGTGTAGTTGCTGCTGGATATGGTTCTGAGAATGGCATGGATTATTGGA 943

Qy    961 TAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAA 1020

Db    944 TTGTTAGGAACTCATGGGGAGCTAAATGGGGAGAGAAGGGCTACCTCAGAGTCCAACGTA 1003

Qy   1021 ACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAAAG 1080

Db   1004 ACGTTGCCAGCTCTAGTGGCTTGTGTGGTTTAGCCACAGAGCCTTCATATCCAGTAAAGA 1063

Qy   1081 AGGAATCAGAAACACTGGTGTCTGCTC 1107

Db   1064 CAGGCGCAAATCCTCCTAAACCCGCTC 1090
```

```
RESULT 9
LEAJ3137
LOCUS       LEAJ3137               1756 bp    mRNA    linear   PLN 07-APR-1998
DEFINITION  Lycopersicon esculentum mRNA for cysteine protease CYP1, complete
            CDS.
ACCESSION   AJ003137
VERSION     AJ003137.1  GI:2828251
KEYWORDS    C14 gene; CYP1; cysteine protease.
SOURCE      Lycopersicon esculentum.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 1756)
  AUTHORS   Lers,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-DEC-1997) Lers A., Dept of Postharvest Science of
            Fresh Produce, The Volcani Center, POB 6, Bet Dagan, 50250, ISRAEL
REFERENCE   2  (bases 1 to 1756)
  AUTHORS   Lers,A., Burd,S., Sonego,L., Khalchitski,A. and Lomaniec,E.
  TITLE     Nucleotide sequence of a full-length C14 cDNA clone (Accession No.
            AJ003137) encoding for cysteine protease CYP1 from tomato
            (PGR98-049)
  JOURNAL   Plant Physiol. 116, 1193-1193 (1998)
FEATURES             Location/Qualifiers
     source          1. .1756
                     /organism="Lycopersicon esculentum"
                     /cultivar="Cherry"
                     /db_xref="taxon:4081"
     gene            29. .1429
                     /gene="C14"
     CDS             29. .1429
                     /gene="C14"
                     /function="cysteine protease"
                     /codon_start=1
                     /product="CYP1"
                     /protein_id="CAA05894.1"
                     /db_xref="GI:2828252"
                     /db_xref="SPTREMBL:O49877"
                     /translation="MAAHSSTLTISILLMLIFSTLSSASDMSIISYDETHIHRRTDDE
                     VSALYESWLIEHGKSYNALGEKDKRFQIFKDNLRYIDEQNSVPNQSYKLGLTKFADLT
                     NEEYRSIYLGTKSSGDRKKLSKNKSDRYLPKVGDSLPESIDWREKGVLVGVKDQGSCG
                     SCWAFSAVAAMESINAIVTGNLISLSEQELVDCDRSYNEGCDGGLMDYAFEFVIKNGG
                     IDTEEDYPYKERNGVCDQYRKNAKVVKIDSYEDVPVNNEKALQKAVAHQPVSIALEAG
                     GRDFQHYKSGIFTGKCGTAVDHGVVIAGYGTENGMDYWIVRNSWGANWGENGYLRVQR
                     NVASSSGLCGLAIEPSYPVKTGPNPPKPAPSPPPSPVKPPTECDEYSQCAVGTTCCCIL
                     QFRRSCFSWGCCPLEGATCCEDHYSCCPHDYPICNVRQGTCSMSKGNPLGVKAMKRIL
                     AQPIGAFGNGGKKSSS"
BASE COUNT      540 a    344 c    406 g    466 t
ORIGIN
```

```
  Query Match             13.8%;  Score 159.4;  DB 8;  Length 1756;
  Best Local Similarity   50.8%;  Pred. No. 1.2e-30;
  Matches  506;  Conservative    0;  Mismatches  461;  Indels   30;  Gaps    4;

Qy    114 ACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTAC 173

Db    149 ACCGACGATGAAGTCTCGGCGTTGTACGAGTCCTGGCTAATCGAGCACGGAAAATCGTAC 208

Qy    174 CATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATC 233

Db    209 AACGCCTTAGGAGAAAAGGATAAGAGATTTCAGATCTTCAAGGATAACTTGAGATACATA 268

Qy    234 AGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTT 293

Db    269 GACGAACAGAACTCTGTTCCGAAT------CAGAGTTATAAGCTCGGATTGACGAAATTC 322

Qy    294 GCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCG 353

Db    323 GCCGATCTGACTAATGAGGAGTACAGGTCGATTTACTTGGGAACGAAGAGTTCCGGTGAC 382

Qy    354 CAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCA 413

Db    383 CGGAAAAAGTTATCGAAGAACAAAAGTGATCGGTATCTTCCTAAAGTTGGGGATAGCTTG 442

Qy    414 CCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGC 473

Db    443 CCGGAATCAATTGACTGGAGAGAAAAAGGTGTGCTTGTTGGTGTCAAGGATCAAGGAAGC 502

Qy    474 TGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCA 533

Db    503 TGTGGGAGTTGTTGGGCATTCTCTGCTGTTGCTGCCATGGAATCAATAAACGCGATAGTC 562

Qy    534 ACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGC 590

Db    563 ACTGGGAATTTGATATCACTATCAGAGCAAGAGTTGGTGGATTGTGATAGGTCGTACAAT 622

Qy    591 GAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGC 650

Db    623 GAAGGTTGCGATGGTGGTCTTATGGACTACGCCTTTGAATTCGTCATTAAGAATGGAGGA 682

Qy    651 ATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAG 710

Db    683 ATCGACACTGAAGAAGACTACCCTTACAAAGAACGCAATGGCGTATGTGATCAATATAGG 742

Qy    711 ATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACA 770

Db    743 AAAAATGCCAAGGTTGTTAA----------------AATAGATAGCTATGAAGATGTT 784

Qy    771 GAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCA 830

Db    785 CCTGTTAATAACGAAAAGGCGTTGCAAAAGGCTGTTGCACATCAACCTGTGAGCATTGCA 844
```

```
Qy    831 ATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACA 890
          |||| ||    |      | ||  |   |      |  | || |      |   | | |
Db    845 CTTGAAGCTGGTGGCAGAGACTTCCAGCAC---TACAAATCTGGTATCTTCACTGGAAAA 901

Qy    891 AGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTA 950
          ||     ||   |  ||||   |||  |  |||  || |||||| | | | ||||  |
Db    902 TGTGGTACTGCAGTGGATCATGGTGTAGTTATTGCTGGATATGGTACTGAGAATGGCATG 961

Qy    951 GATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGG 1010
          ||||| ||||| |  |  || ||||||||||   | ||||||||  | || ||| |  |
Db    962 GATTATTGGATCGTTAGGAACTCATGGGGAGCTAACTGGGGAGAGAACGGCTACCTCAGA 1021

Qy   1011 ATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATAC 1070
          ||||  | |||   |  | |     || |||||||  |           |||||||
Db   1022 GTCCAGCGTAACGTTGCCAGCTCTAGTGGCTTGTGTGGTTTAGCCATAGAGCCTTCATAT 1081

Qy   1071 CCAACCAAAGAGGAATCAGAAACACTGGTGTCTGCTC 1107
          |||   |||   | | || |  | |       | ||||
Db   1082 CCAGTAAAAACAGGACCAAATCCTCCTAAACCCGCTC 1118
```

RESULT 10
AF172856

```
LOCUS       AF172856               1779 bp    mRNA    linear   PLN 17-AUG-2001
DEFINITION  Lycopersicon esculentum cysteine protease TDI-65 (tdi-65) mRNA,
            complete cds.
ACCESSION   AF172856
VERSION     AF172856.1  GI:5726640
KEYWORDS    .
SOURCE      Lycopersicon esculentum.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 1779)
  AUTHORS   Harrak,H., Azelmat,S., Baker,E.N. and Tabaeizadeh,Z.
  TITLE     Isolation and characterization of a gene encoding a drought-induced
            cysteine protease in tomato (Lycopersicon esculentum)
  JOURNAL   Genome 44 (3), 368-374 (2001)
  MEDLINE   21337597
   PUBMED   11444695
REFERENCE   2  (bases 1 to 1779)
  AUTHORS   Harrak,H. and Tabaeizadeh,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1999) Biological Sciences, University of Quebec
            in Montreal, P.O. Box 8888 Station 'Centre Ville', Montreal, Quebec
            H3C 3P8, Canada
FEATURES             Location/Qualifiers
     source          1. .1779
                     /organism="Lycopersicon esculentum"
                     /db_xref="taxon:4081"
     gene            1. .1779
                     /gene="tdi-65"
     CDS             48. .1448
                     /gene="tdi-65"
                     /note="induced by drought; localized in the nuclei and
                     chloroplast (Tabaeizadeh, Z. et al., 1995. Protoplasma,
                     186:208-219)"
                     /codon_start=1
                     /product="cysteine protease TDI-65"
                     /protein_id="AAD48496.1"
                     /db_xref="GI:5726641"
                     /translation="MAAHSSTLTISILLMLIFSTLSSASDMSIISYDETHIHRRTDDE
                     VSALYESWLIEHGKSYNALGEKDKRFQIFKDNLRYIDEQNSVPNQSYKLGLTKFADLT
                     NEEYRSIYLGTKSSGDRKKLSKNKSDRYLPKVGDSLPESIDWREKGVLVGVKDQGSCG
                     SCWAFSAVAAMESINAIVTGNLISLSEQELVDCDRSYNEGCDGGLMDYAFEFVIKNGG
                     IDTEEDYPYKERNGVCDQYRKNAKVVKIDSYEDVPVNNEKALQKAVAHQPVSIALEAG
                     GRDFQHYKSGIFTGKCGTAVDHGVVIAGYGTENGMDYWIVRNSWGANWGENGYLRVQR
                     NVASSSGLCGLAIEPSYPVKTGPNPPKPAPSPPSPVKPPTECDEYSQCAVGTTCCCIL
                     QFRRSCFSWGCCPLEGATCCEDHYSCCPHDYPICNVRQGTCSMSKGNPLGVKAMKRIL
                     AQPIGAFGNGGKKSSS"
```

```
BASE COUNT      543 a    356 c    410 g    470 t
ORIGIN

  Query Match           13.8%;  Score 159.4;  DB 8;  Length 1779;
  Best Local Similarity  50.8%;  Pred. No. 1.2e-30;
  Matches  506;  Conservative    0;  Mismatches  461;  Indels   30;  Gaps    4;

Qy    114 ACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTAC 173
          ||  |  |  | || ||  |  ||| | |    |||   |  ||||| |||      |||
Db    168 ACCGACGATGAAGTCTCGGCGTTGTACGAGTCCTGGCTAATCGAGCACGGAAAATCGTAC 227

Qy    174 CATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATC 233
          |   |   | |||| |||  |||||| || |||| |||||| |||||| ||  || ||
Db    228 AACGCCTTAGGAGAAAAGGATAAGAGATTTCAGATCTTCAAGGATAACTTGAGATACATA 287

Qy    234 AGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTT 293
             ||  |||  |       ||       ||   | ||    | |||||||  || ||
Db    288 GACGAACAGAACTCTGTTCCGAAT------CAGAGTTATAAGCTCGGATTGACGAAATTC 341

Qy    294 GCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCG 353
          || ||  | |||  | | |||| |||    |  ||||||  | |    |     |
Db    342 GCCGATCTGACTAATGAGGAGTACAGGTCGATTTACTTGGGAACGAAGAGTTCCGGTGAC 401

Qy    354 CAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCA 413
          | | |||    | |  |||||  |       | |    |   |   | | ||
Db    402 CGGAAAAAGTTATCGAAGAACAAAAGTGATCGGTATCTTCCTAAAGTTGGGGATAGCTTG 461

Qy    414 CCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGC 473
          || | ||||   || |||||  ||||||||||  |       || ||| | |||||  ||
Db    462 CCGGAATCAATTGACTGGAGAGAAAAAGGTGTGCTTGTTGGTGTCAAGGATCAAGGAAGC 521

Qy    474 TGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCA 533
          |||||  ||  ||||||| || |||||    |  ||||| ||| ||  | | || ||||
Db    522 TGTGGGAGTTGTTGGGCATTCTCTGCTGTTGCTGCCATGGAATCAATAAACGCGATAGTC 581

Qy    534 ACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGC 590
          || ||  |  |  |    || || || |||||  | || || ||||    |    | |
Db    582 ACTGGGAATTTGATATCACTATCAGAGCAAGAGTTGGTGGATTGTGATAGGTCGTACAAT 641

Qy    591 GAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGC 650
          |||||||   |  |||     | |  |  | || ||||  ||  |  |  |||| ||
Db    642 GAAGGTTGCGATGGTGGTCTTATGGACTACGCCTTTGAATTCGTCATTAAGAATGGAGGA 701

Qy    651 ATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAG 710
          || | |||||| || || || ||||||| ||       | ||   |||  |     ||| |
Db    702 ATCGACACTGAAGAAGACTACCCTTACAAAGAACGCAATGGCGTATGTGATCAATATAGG 761

Qy    711 ATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACA 770
          | | | ||||||| |                    ||||       ||||   |
Db    762 AAAAATGCCAAGGTTGTTAA----------------AATAGATAGCTATGAAGATGTT 803

Qy    771 GAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCA 830
                | |  ||  | |||||   ||  ||  |     | |||||  | ||  |  ||
Db    804 CCTGTTAATAACGAAAAGGCGTTGCAAAAGGCTGTTGCACATCAACCTGTGAGCATTGCA 863

Qy    831 ATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACA 890
          ||||  ||    |      | ||  |   |      |  | || |      |   | | |
Db    864 CTTGAAGCTGGTGGCAGAGACTTCCAGCAC---TACAAATCTGGTATCTTCACTGGAAAA 920

Qy    891 AGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTA 950
          ||     ||   |  ||||   |||  |  |||  || |||||| | | | ||||  |
Db    921 TGTGGTACTGCAGTGGATCATGGTGTAGTTATTGCTGGATATGGTACTGAGAATGGCATG 980

Qy    951 GATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGG 1010
          ||||| ||||| |  |  || ||||||||||  | ||||||||  | || ||| |  |
Db    981 GATTATTGGATCGTTAGGAACTCATGGGGAGCTAACTGGGGAGAGAACGGCTACCTCAGA 1040

Qy   1011 ATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATAC 1070
          ||||  | |||   |  | |     || |||||||  |           |||||||
Db   1041 GTCCAGCGTAACGTTGCCAGCTCTAGTGGCTTGTGTGGTTTAGCCATAGAGCCTTCATAT 1100

Qy   1071 CCAACCAAAGAGGAATCAGAAACACTGGTGTCTGCTC 1107
          |||   |||   | | || |  | |       | ||||
```

```
Db   1101 CCAGTAAAAACAGGACCAAATCCTCCTAAACCCGCTC 1137


RESULT 11
AF259983
LOCUS       AF259983               1751 bp    mRNA    linear   PLN 01-MAY-2001
DEFINITION  Ipomoea batatas putative cysteine protease mRNA, complete cds.
ACCESSION   AF259983
VERSION     AF259983.1  GI:13897889
KEYWORDS    .
SOURCE      Ipomoea batatas.
  ORGANISM  Ipomoea batatas
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
REFERENCE   1  (bases 1 to 1751)
  AUTHORS   Chen,H.-J., Hou,W.-C. and Lin,Y.-H.
  TITLE     Isolation of a cDNA clone encoding a putative cysteine protease
            from senescent leaves of sweet potato (Ipomoea batatas cv. Tainong
            57)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1751)
  AUTHORS   Chen,H.-J., Hou,W.-C. and Lin,Y.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2000) Academia Sinica, Institute of Botany,
            Nankang, Taipei, Taiwan 11529, Republic of China
FEATURES             Location/Qualifiers
     source          1. .1751
                     /organism="Ipomoea batatas"
                     /cultivar="Tainong 57"
                     /db_xref="taxon:4120"
                     /tissue_type="senescent leaf"
     CDS             28. .1416
                     /codon_start=1
                     /product="putative cysteine protease"
                     /protein_id="AAK48495.1"
                     /db_xref="GI:13897890"
                     /translation="MAKLLILSLFVLAAVSSASASADMSIITYDEEHPAKGLSRSDEE
                     VMALYESWLVEHGKSYNGLGGEKDKRFEIFKDNLRYIDEQNSRGDRSYKLGLNRFADL
                     TNEEYRSTYLGAKTDARRRIAKTKSDRRYAPKAGGSLPDSIDWREKGAVAEVKDQGSC
                     GSCWAFSTIAAVEGINQIVTGELISLSEQELVDCDTSYNEGCNGGLMDYAFEFIIKNG
                     GIDTEADYPYTGRYGRCDQTRKNAKVVSIDGYEDVTPYDEAALKEAVAGQPVSVAIEA
                     GGRDFQLYSSGIFTGSCGTDLDHGVTAVGYGTENGVDYWIVKNSWAASWGEKGYLRMQ
                     RNVKDKNGLCGIAIEPSYPTKTGENPPNPGPSPPSPVSPPNMCDDYDECPTSTTCCCV
                     FPYGEHCFAWGCSPLESAVCCEDHYSCCPHDYPVCHVSQGTCPMSKNSPLGVKPMRRT
                     PAKKIRNNGSSS"
BASE COUNT      455 a    397 c    460 g    439 t
ORIGIN

  Query Match           13.4%;  Score 155.4;  DB 8;  Length 1751;
  Best Local Similarity   52.7%;  Pred. No. 1.4e-29;
  Matches  476;  Conservative    0;  Mismatches  391;  Indels   36;  Gaps    5;

Qy     182 CGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACAT 241
           || |||  |||  |||||  ||||||| |||||| ||||    |  || |||   ||
Db     219 CGGAGAGAAGGATAAGAGGTTTGAGATCTTCAAGGATAATCTCAGGTACATCGACGAGCA 278

Qy     242 GAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACAT 301
           |      |||||      |  ||  ||  |     | ||| ||||||  || |||||  |
Db     279 G------AACAGCCGGGGAGACCGGTCCTACAAGCTCGGACTGAACAGATTCGCTGATCT 332

Qy     302 CACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAAT 361
           |||     | |||| | |   || |||| |    |||   |  ||||     | |  |||
Db     333 CACCAACGAGGAGTACCGGTCAACGTACCTCGGCGCTAAGACCGATGCTCGCCGGAGAAT 392

Qy     362 CAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATC 421
           |   | | | || |   |   |  | | |  |   |  | | |      || |  ||
Db     393 CGCTAAGACGAAAAGCGATCGGCGGTATGCTCCCAAGGCCGGCGGCAGCTTGCCGGACTC 452

Qy     422 ATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAG 481
             || |||||  | |||||| | || || |||| ||  | |||||  |||||| ||
Db     453 CATTGACTGGAGAGAGAAAGGTGCCGTCGCCGAAGTCAAAGATCAAGGAAGCTGTGGGAG 512

Qy     482 GGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGA 541
           |  |||||| || ||  |||  |  ||  | ||||     | |   ||| |  || || ||
Db     513 TTGCTGGGCATTCTCAACCATTGCTGCTGTGGAAGGTATAAACCAAATCGTGACTGGTGA 572

Qy     542 CCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGCGAAGGTTC 598
           |  |      | |||||||| || || || || ||||       | | | ||||||  |
Db     573 ATTGATCTCATTGTCTGAACAGGAGCTTGTGGATTGTGATACATCATATAACGAAGGCTG 632

Qy     599 TTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCAC 658
           |   |||      | |  |  | |||||  |   |  |  |  ||||||| ||||  ||
Db     633 CAATGGTGGTCTAATGGACTATGCCTTCGAGTTCATCATCAAGAATGGTGGTATTGATAC 692

Qy     659 TGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGA 718
           ||| |  ||||| || ||||  | ||   | || || ||  |    |  | || | | |
Db     693 TGAGGCAGATTACCCCTACACTGGTAGGTATGGAAGGTGTGATCAAACAAGGAAAAATGC 752

Qy     719 CAAG---GTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATC 775
           |||   ||| | ||||| || |||||   | | |      |  ||||||    ||  |
Db     753 TAAGGTCGTTTCTATTGATGGGTATGAGGATGTTACTCCTTATGATGAGGCGGCACTAAA 812

Qy     776 AGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGA 835
           |||| |||                 ||   ||           ||    |   |  ||
Db     813 AGAGGCAGT---------------TGCAGGCCAACCAGTTAGTGTTGCCATTGAAGCTGG 857

Qy     836 TGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCC 895
           ||  | ||| || ||  | ||  |   ||| || | |  |||| |  |    | |||
Db     858 TGGCAGAGACTTCCAACTCTATTCATCGGGTATATTTACTGGATCATGTGGAACA----- 912

Qy     896 GTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTA 955
           |   |  | |||  |||      ||| || |||||| |||   ||||||||||||
Db     913 ----GACCTGGACCACGGTGTGACTGCTGTTGGATATGGTACAGAAAATGGTGTAGATTA 968

Qy     956 CTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCA 1015
           |||||  |  || || |||||||  | |   |||||||  | || ||  |  |||| ||
Db     969 TTGGATTGTTAAGAACTCATGGGCTGCATCCTGGGGAGAGAAAGGCTATTTGAGGATGCA 1028

Qy    1016 AAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAAC 1075
           | ||   |    |   |   ||  |||||| ||          |||| ||||| ||
Db    1029 GCGTAATGTGAAAGACAAAAATGGGTTGTGTGGTATCGCGATAGAGCCTTCTTACCCGAC 1088

Qy    1076 CAA 1078
           |||
Db    1089 CAA 1091


RESULT 12
AF454957
LOCUS       AF454957               1699 bp    mRNA    linear   PLN 14-JAN-2002
DEFINITION  Brassica oleracea senescence-associated cysteine protease (CP2)
            mRNA, complete cds.
ACCESSION   AF454957
VERSION     AF454957.1  GI:18141282
KEYWORDS    .
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 1699)
  AUTHORS   Coupe,S.A., Sinclair,B.K., Watson,L.M. and Eason,J.R.
  TITLE     Cysteine proteases and broccoli senescence: cloning,
            characterization and contribution to the process
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1699)
  AUTHORS   Coupe,S.A., Sinclair,B.K., Watson,L.M. and Eason,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Crop and Food Research, Private Bag 11 600,
            Palmerston North 5301, New Zealand
FEATURES             Location/Qualifiers
     source          1. .1699
                     /organism="Brassica oleracea"
                     /db_xref="taxon:3712"
     gene            1. .1699
```

```
                     /gene="CP2"
     CDS             25. .1407
                     /gene="CP2"
                     /note="BoCP2; similar to Arabidopsis thaliana cysteine
                     protease component of a protease-inhibitor complex;
                     expressed during senescence"
                     /codon_start=1
                     /product="senescence-associated cysteine protease"
                     /protein_id="AAL60579.1"
                     /db_xref="GI:18141283"
                     /translation="MGSVKVTILLLAMMIGVSYAADMSIISYDEKHHITAENERSDAE
                     VARIYEAWMEKHGKKAQSNGLVGEEKDQRFEIFKDNLRFIDEHNNKNLSYKLGLTRFA
                     DLTNEEYRSIYLGAKSKKRVLKTSDRYQPRVGDAIPDSVDWRKEGAVAAVKDQGSCGS
                     CWAFSTIGAVEGINKIVTGDLISLSEQELVDCDTSYNQGCNGGLMDYAFEFIIKNGGI
                     DTEEDYPYKAADGRCDQTRKNAKVVTIDAYEDVPENNEAALKKTLANQPISVAIEAGG
                     RAFQLYSSGVFDGICGTELDHGVVAVGYGTENGKDYWIVRNSWGGSWGESGYIKMARN
                     IAEPTGKCGIAMEASYPIKKGQNPPNPGPSPPSPIKPPTQCDKYYSCPESNTCCCLFK
                     YGKYCFGWGCCPLEAATCCDDNTSCCPHEYPVCNGDTCLMSKNSPFSVKALKRTPAKP
                     FWAHSRKNIA"
BASE COUNT      491 a    359 c    421 g    428 t
ORIGIN

  Query Match             13.4%;  Score 155.2;  DB 8;  Length 1699;
  Best Local Similarity   55.2%;  Pred. No. 1.5e-29;
  Matches  379;  Conservative    0;  Mismatches  283;  Indels   24;  Gaps    3;

Qy    401 TTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAA 460
          | |||||     |||||  |||   || ||||| ||| |||||| | || ||   || ||
Db    423 TGGTGACGCGATACCTGATTCAGTTGACTGGAGAAAAGAAGGTGCCGTCGCCGCCGTTAA 482

Qy    461 GTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGC 520
          | || ||  |||| || ||  |||||||| |||||  |||  |||||  | |||| |
Db    483 AGATCAGGGAAGCTGCGGGAGTTGTTGGGCATTTTCAACCATTGGAGCGGTGGAAGGAAT 542

Qy    521 ACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGT 580
          | |    || |  |||||||||| |  |      | |||||||||||||| || |||||||
Db    543 AAACAAGATCGTGACAGGAGACTTAATCTCTTTGTCTGAACAAGAACTTGTTGACTGTGA 602

Qy    581 GGAA---GAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTT 637
             |    | | | | |||| | |    || | |  | || |  |||| ||||    |  |
Db    603 CACATCTTACAACCAGGGTTGTAATGGAGGTTTGATGGATTATGCGTTTGAATTCATCAT 662

Qy    638 AGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATG 697
          || | ||||| |||  || || || |||||||| |||| ||||    || ||| | ||
Db    663 TAAAAACGGTGGTATTGATACCGAAGAGGATTATCCATACAAAGCTGCTGACGGTCGTTG 722

Qy    698 CAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTC 757
          |    |  | || | | |  ||||||   |                    |||     |
Db    723 TGATCAGACCAGGAAAAACGCAAAGGTTGTCA------------------CAATTGATGC 764

Qy    758 AGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACC 817
          |||||  |          | |  |||   ||| ||   |   |   |  | | || ||
Db    765 CTATGAGGATGTGCCTGAGAACAACGAGGCTGCGCTCAAGAAAACTTTGGCTAACCAGCC 824

Qy    818 AATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGG 877
          |||||||||  | || || ||    | |       ||  |   |         ||  | ||
Db    825 TATTAGTGTCGCCATCGAAGCTGGGGGTCGGGCATTCCAGCTC---TATTCTTCGGGTGT 881

Qy    878 AGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTC 937
          |    |  | | | ||       | | |  | ||    ||| ||  |||||||| ||  |
Db    882 ATTTGATGGAATATGTGGAACAGAGCTAGACCATGGAGTTGTAGCTGTGGGTTACGGAAC 941

Qy    938 AGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA 997
          | | | ||    || |||||||| | || ||| ||||||||||  |   |||||||
Db    942 TGAGAACGGAAAGGACTACTGGATTGTGAGAAACTCATGGGGAGGAAGCTGGGGAGAGAG 1001

Qy    998 TGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTA 1057
           || |||||   |||   | | ||||  | | |    |  ||||  ||||||  ||
Db   1002 CGGATACATAAAGATGGCACGTAACATTGCTGAGCCAACAGGAAAGTGTGGAATCGCCAT 1061

Qy   1058 TTTCGCTTCATACCCAACCAAAGAGG 1083
             |||||||||||  | |||  | |
Db   1062 GGAGGCTTCATACCCGATCAAGAAAG 1087
```

```
RESULT 13
AR049578
LOCUS       AR049578               1056 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5824534.
ACCESSION   AR049578
VERSION     AR049578.1  GI:6005617
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Asano,M., Kawai,M., Miwa,T. and Nio,N.
  TITLE     Aminopeptidase GX, and a method of hydrolyzing a protein with the
            same
  JOURNAL   Patent: US 5824534-A 1 20-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="unknown"
BASE COUNT      295 a    229 c    297 g    235 t
ORIGIN

  Query Match             13.4%;  Score 154.4;  DB 6;  Length 1056;
  Best Local Similarity   52.9%;  Pred. No. 2.6e-29;
  Matches  494;  Conservative    0;  Mismatches  401;  Indels   39;  Gaps    6;

Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
          |||  |||   |||  || |  || ||   ||  ||| | | |  |||        |||| |
Db     38 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 97

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
          |    || ||| |    | ||  ||  |||  ||| | |  |||| |||||| | |||
Db     98 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 157

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
          |   |  |||   |||   ||  |        |   | || |    |       | |||
Db    158 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 211

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
          |||||  ||| |||||  ||||      | || | |||    |||||||||  | |    |
Db    212 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 271

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
          |||       | |    ||  || | ||   |  ||      |   |  ||   |
Db    272 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 328

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
          | ||| |    ||||  ||    ||||||||||  ||||||   |  | |  || ||
Db    329 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 388

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
          |||||||  |||||||| ||  ||||||| || || || |  || ||  ||||||  |  |
Db    389 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 448

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
          ||   |||| |||||| || ||  ||     | || |||||||| | || || ||||
Db    449 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 508

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
          | | | | | |||| |   |     |||      | |  |  | || || |   |  |
Db    509 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 568

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
          |  |||| |||||||   |||||| ||||| || ||| | | |    || |||||||||
Db    569 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 628

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756
          |  |  ||| || | | |  ||    ||| | ||||| |  || |||  | |       |
Db    629 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 688

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
          ||||||                        ||| ||   ||  ||| |     | || |
```

```
Db    689 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 727

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
          |  | || ||  | ||||| | |   |      | ||  |        |   | | || ||
Db    728 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 784

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
          |   |   | | | ||      |   |  |||| |||  |   ||| || |||||
Db    785 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 844

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
          ||||  | |||   ||||| ||||| |  |  ||||||||||        |||||||| |
Db    845 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 904

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
          |||| |||||  |  |  |||||||    | |||
Db    905 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 938


RESULT 14
E13053
LOCUS       E13053                  1056 bp    DNA     linear   PAT 27-APR-1998
DEFINITION  Modified cDNA encoding thiol protease D3-beta.
ACCESSION   E13053
VERSION     E13053.1  GI:3251865
KEYWORDS    JP 1997121870-A/3.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Kawai,M., Asano,M., Suzuki,S., Miwa,T., Shibai,H. and Yokozeki,K.
  TITLE     DNA CAPABLE OF CODING NEW THIOL PROTEASE AND PRODUCTION OF THE SAME
            THIOL PROTEASE USING THE DNA
  JOURNAL   Patent: JP 1997121870-A 3 13-MAY-1997;
            AJINOMOTO CO INC
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1997121870-A/3
            PD   13-MAY-1997
            PF   28-DEC-1995 JP 1995353931
            PR   29-DEC-1994 JP 94P   340399, 30-AUG-1995 JP 95P   245279 PI
            KAWAI MISAKO, ASANO MINAO, SUZUKI SHUNICHI, MIWA TETSUYA, PI
            SHIBAI HIROSHIRO, YOKOZEKI KENZO
            PC   C12N15/09,C12N1/19,C12N1/21,C12N9/50,(C12N1/19,C12R1:865), PC
             (C12N1/21,
            PC   C12R1:19),(C12N9/50,C12R1:19),(C12N9/50,C12R1:865); CC
            strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1056
            FT                   /organism='Artificial sequences' FT   CDS
                   1. .1056
            FT                   /product='modified thiol protease D3-beta'.
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      295 a    229 c    297 g    235 t
ORIGIN

  Query Match             13.4%;  Score 154.4;  DB 6;  Length 1056;
  Best Local Similarity   52.9%;  Pred. No. 2.6e-29;
  Matches  494;  Conservative    0;  Mismatches  401;  Indels   39;  Gaps    6;

Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
          |||  |||   |||  || |  || ||   ||  ||| | | |  |||       |||| |
Db     38 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 97

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
          |    || ||| |    | ||  ||  |||  ||| |  |  |||| |||||| | |||

Db     98 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 157

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
          |   |  |||   |||   ||  |          |   | || |    |       | |||
Db    158 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 211

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
          |||||  ||| |||||  ||||     | || | |||    |||||||||  | |    |
Db    212 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 271

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
            |||       | |    ||  || | ||   |  ||      |   |  ||    |
Db    272 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 328

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
          | ||| |    |||  ||    ||||||||||  ||||||   |  | |  || ||
Db    329 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 388

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
          |||||||  |||||||| ||  |||||||  || || || |  || ||  ||||||  |  |
Db    389 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 448

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
          ||   |||| |||||| || ||  ||    | || |||||||||  | || || ||||
Db    449 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 508

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
          | | | | | |||| |    |    |||    | |  |  | || || |   |  |
Db    509 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 568

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
          |  |||| |||||||   |||||||  ||||| || ||| | | |    || |||||||||
Db    569 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 628

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756
          |  |  ||| || | | |  ||    ||| | ||||| |  || |||  | |       |
Db    629 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 688

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
          ||||||                    ||| ||   ||  ||| |     | || |
Db    689 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 727

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
          |  | || ||  | ||||| | |   |      | ||  |        |   | | || ||
Db    728 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 784

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
          |   |   | | | ||      |   |  |||| |||  |   ||| || |||||
Db    785 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 844

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
          ||||  | |||   ||||| ||||| |  |  ||||||||||        |||||||| |
Db    845 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 904

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
          |||| |||||  |  |  |||||||    | |||
Db    905 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 938


RESULT 15
E14274
LOCUS       E14274                  1056 bp    DNA     linear   PAT 28-JUL-1999
DEFINITION  DNA encoding Glycine max protease D3.
ACCESSION   E14274
VERSION     E14274.1  GI:5708957
KEYWORDS    JP 1997294583-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Asano,M., Kawai,M., Miwa,T. and Nio,N.
  TITLE     AMINOPEPTIDASE GX AND HYDROLYSIS OF PROTEIN USING THE SAME
  JOURNAL   Patent: JP 1997294583-A 1 18-NOV-1997;
```

```
            AJINOMOTO CO INC
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1997294583-A/1
            PD   18-NOV-1997
            PF   14-FEB-1997 JP 1997030458
            PR   08-MAR-1996 JP 96P    51848
            PI   ASANO MINAO, KAWAI MISAKO, MIWA TETSUYA, NIO NORIKI PC
            C12N9/48//C12N15/00;
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1056
            FT                   /organism='Artificial gene'
            FT   CDS             1. .1056
            FT                   /product='Glycine max protease D3' FT
            misc_feature    1. .9
            FT                   /note='vector-derived sequence'.
FEATURES              Location/Qualifiers
     source           1. .1056
                      /organism="unidentified"
                      /db_xref="taxon:32644"
BASE COUNT      295 a    229 c    297 g    235 t
ORIGIN

  Query Match             13.4%;  Score 154.4;  DB 6;  Length 1056;
  Best Local Similarity   52.9%;  Pred. No. 2.6e-29;
  Matches  494;  Conservative    0;  Mismatches  401;  Indels   39;  Gaps    6;

Qy     103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
           ||| |||   |||  || |  || ||   ||  ||| | | |  |||       |||| |
Db      38 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 97

Qy     163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
           |    || ||| |    | ||  ||  |||  ||| |  |  |||| |||||| | |||
Db      98 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 157

Qy     223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
           |  |  |||  |||  ||  |       |  | || |    |      | |||
Db     158 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 211

Qy     283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
           ||||| ||| |||||  ||||     | || | |||    |||||||||  | |    |
Db     212 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 271

Qy     343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
           |||      | |    ||  || | ||   |  ||     |   |  ||    |
Db     272 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 328

Qy     403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
           | ||| |    ||||  ||    ||||||||||| ||||||   |  | |  || ||
Db     329 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 388

Qy     463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
           ||||||| |||||||| ||  ||||||| || || || |  || ||  ||||||  |  |
Db     389 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 448

Qy     523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
           ||   |||| |||||| || ||  ||     | || |||||||| | || || ||||
Db     449 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 508

Qy     580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
           | | | | |||| |    |    |||     | |  |  | || || |   |  |
Db     509 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 568

Qy     640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
           |  |||| |||||||   ||||||| ||||| || ||| | | |   || |||||||||
Db     569 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 628

Qy     700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756
           |  |  ||| || | | |  ||    ||| | ||||| |  || |||  | |      |
Db     629 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 688

Qy     757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
             ||||||                    ||| ||   ||  ||| |     | || |
Db     689 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 727

Qy     817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
           |  | || ||  | ||||| | |   |       | ||  |       |   | | || ||
Db     728 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 784

Qy     877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
           |   |   | | | ||       |   |  ||||  |||  |  ||| || |||||
Db     785 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 844

Qy     937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
           ||||  | |||   ||||| ||||| |  |  ||||||||||       |||||||| |
Db     845 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 904

Qy     997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
           |||| |||||  |  |  |||||||    | |||
Db     905 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 938


Search completed: January 11, 2003, 20:37:37
Job time : 2682 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 19:45:12 ; Search time 1928 Seconds
(without alignments)
9710.572 Million cell updates/sec

Title: US-09-805-694B-1
Perfect score: 1156
Sequence: 1 gcggccgcatgggtttcctt.....tctcctctttgagcggccgc 1156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	775	67.0	868	10	BE660305	BE660305 483 GmaxS
2	700	60.6	845	10	BE660308	BE660308 1121 Gmax
3	692	59.9	756	9	AI736102	AI736102 sb22g10.y
4	690.2	59.7	729	9	AI941099	AI941099 sb84d09.y
5	662.4	57.3	732	9	AI941069	AI941069 sb84a04.y
6	659.8	57.1	705	9	AI748608	AI748608 sb59e03.y
7	655.8	56.7	704	10	AW472063	AW472063 si19e07.y
8	649.8	56.2	664	10	BE660304	BE660304 2-C10 Gma
9	648.4	56.1	659	10	AW317682	AW317682 sg56a06.y
10	646	55.9	657	10	AW317681	AW317681 sg56a05.y
11	635.4	55.0	655	10	AW433000	AW433000 si03b11.y
12	633.6	54.8	657	10	AW398014	AW398014 sg71b10.y
13	630	54.5	665	9	AI794978	AI794978 sb74e04.y
14	626.4	54.2	654	10	AW397984	AW397984 sg70g06.y
15	625	54.1	677	10	AW318029	AW318029 sg60c09.y
16	624.8	54.0	652	10	AW395647	AW395647 sg73e11.y
17	622.6	53.9	673	9	AI748618	AI748618 sb59f03.y
18	622.2	53.8	721	9	AI794998	AI794998 sb74h04.y
19	619.8	53.6	639	10	AW317968	AW317968 sg59e05.y
20	619	53.5	695	10	AW472076	AW472076 si19f08.y
21	615.6	53.3	707	12	BG156896	BG156896 sab21c10.
22	615	53.2	662	10	AW472405	AW472405 si24g02.y
23	614	53.1	656	10	AW317722	AW317722 sg56e01.y
24	613	53.0	621	10	AW397343	AW397343 sg77f10.y
c 25	612.4	53.0	806	10	BE657469	BE657469 GM700001B
26	611.2	52.9	624	10	AW318230	AW318230 sg62f06.y
27	609.8	52.8	630	10	AW397481	AW397481 sg79d08.y
28	609.8	52.8	666	10	AW398031	AW398031 sg71d04.y
29	607.6	52.6	644	10	AW318267	AW318267 sg63a09.y
30	607.4	52.5	633	9	AI795020	AI795020 sb75b09.y
31	605.4	52.4	615	10	AW397146	AW397146 sg67d12.y
32	604.2	52.3	660	10	AW318176	AW318176 sg62a08.y
33	602.4	52.1	637	10	AW432985	AW432985 si03a07.y
34	600.4	51.9	627	10	AW397002	AW397002 sg65g10.y
35	597.2	51.7	650	10	AW397239	AW397239 sg76e03.y
36	595.6	51.5	635	10	AW397175	AW397175 sg67g06.y
37	595.4	51.5	627	10	AW317822	AW317822 sg57f08.y
38	594.8	51.5	612	9	AI736014	AI736014 sb21g10.y
39	593.4	51.3	644	10	AW318150	AW318150 sg61a10.y
40	592.6	51.3	617	10	AW396990	AW396990 sg65f08.y
41	590.8	51.1	608	10	AW760065	AW760065 si57e11.y
42	587.6	50.8	609	9	AI941141	AI941141 sb85b01.y
43	587.6	50.8	623	10	AW396995	AW396995 sg65g01.y
44	587.2	50.8	601	10	AW397310	AW397310 sg77c10.y
45	585	50.6	602	10	AW397972	AW397972 sg70f06.y

ALIGNMENTS

RESULT 1
BE660305
LOCUS BE660305 868 bp mRNA linear EST 06-SEP-2000
DEFINITION 483 GmaxSC Glycine max cDNA, mRNA sequence.
ACCESSION BE660305
VERSION BE660305.1 GI:9986197
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 868)
AUTHORS Harris,N., Chapman,B.P. and Gijzen,M.
TITLE Gene expression in developing soybean seed coats
JOURNAL Unpublished (2000)
COMMENT Contact: Gijzen M
Agriculture and Agri-Food Canada
1391 Sandford Street, London, Ontario, Canada N5V 4T3
Tel: 519 457 1470
Fax: 519 457 3997
Email: gijzenm@em.agr.ca.
FEATURES Location/Qualifiers
source 1. .868
/organism="Glycine max"
/cultivar="Harosoy 63"
/db_xref="taxon:3847"
/clone_lib="GmaxSC"

/tissue_type="Seed coats"
/lab_host="E. coli strain XLOLR"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from polyA+ enriched mRNA from green seed coats in mid to late developemental stage , average fresh weight 250 mg per seed. Traces of pod and embryo tissue also present. Complementary DNA was synthesized from mRNA using an XhoI-poly(dT) linker-primer. EcoRI adopters were ligated to the blunt-ended cDNA fragments and the products were digested with XhoI for directional cloning into lambda ZAP Express vector. This lambda library was amplified once using E. coli host strain XL1 Blue MRF'. Inserts were then subcloned by mass excision using ExAssist helper phage for conversion into phagemid vector pBK-CMV in E. coli host strain XLOLR."

BASE COUNT 278 a 177 c 192 g 214 t 7 others
ORIGIN

Query Match 67.0%; Score 775; DB 10; Length 868;
Best Local Similarity 95.3%; Pred. No. 9.3e-207;
Matches 827; Conservative 0; Mismatches 37; Indels 4; Gaps 3;

```
Qy     29 TCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATATCAACTCATCGTTCCATATT 88
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATATCAACTCATCGTTCCATATT 60

Qy     89 GGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATG 148
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATG 120

Qy    149 GAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGAT 208
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGAT 180

Qy    209 TTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTC 268
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTC 240

Qy    269 TCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTA 328
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTA 300

Qy    329 CTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAA 388
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAA 360

Qy    389 GGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCAT 448
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCAT 420

Qy    449 CACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGC 508
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    421 CACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGGGCGTTTTCTGCCACGGGAGC 480

Qy    509 CATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACT 568
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACT 540

Qy    569 CGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGA 628
          |||||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||
Db    541 CGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGATGGCACTATCAATCGTTCGA 600

Qy    629 ATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGA 688
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    601 ATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGA 660

Qy    689 GGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCT 748
          |||||||||||| |||||||| |||||||||||||||||||||||| ||||||| |
Db    661 NGGTAGATGCAAAGNCAATAAGATCCAAGACAAGGTTACAATTGACGGATTTGAAACTTT 720

Qy    749 AATAATGTCAGA-TGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCA-TC 806
          ||||||||| | ||||||| | ||||||||| | || ||||||||||||| ||||| ||
```

```
Db    721 TATAATGTCANATTGAGAGTCCGGAATCAGAGNCNGAACAAGCGTTCTTAAACGCCATTC 780

Qy    807 CTTGAGCAACCAATTAGTGTCT--CAATTGATGCAAAAGATTTTCATTTATACACCGGGG 864
          |||||||||||||| ||| | ||  || |     ||||||| |||||||| | | ||||||
Db    781 CTTGAGCAACCAAATAGNGGCTTCCATTGATGCCAAAAGAATTTCATTTTTNCCCCGGGG 840

Qy    865 GAATTTATGATGGAGAAAACTGTACAAG 892
          |||||  |||||| ||||||| | | ||
Db    841 GAATTATTGATGGGGAAAACTTTTCCAG 868


RESULT 2
BE660308
LOCUS       BE660308                845 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  1121 GmaxSC Glycine max cDNA, mRNA sequence.
ACCESSION   BE660308
VERSION     BE660308.1  GI:9986200
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Harris,N., Chapman,B.P. and Gijzen,M.
  TITLE     Gene expression in developing soybean seed coats
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Gijzen M
            Agriculture and Agri-Food Canada
            1391 Sandford Street, London, Ontario, Canada N5V 4T3
            Tel: 519 457 1470
            Fax: 519 457 3997
            Email: gijzenm@em.agr.ca.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Glycine max"
                     /cultivar="Harosoy 63"
                     /db_xref="taxon:3847"
                     /clone_lib="GmaxSC"
                     /tissue_type="Seed coats"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
                     cDNA library was constructed from polyA+ enriched mRNA
                     from green seed coats in mid to late developemental stage
                     , average fresh weight 250 mg per seed. Traces of pod and
                     embryo tissue also present. Complementary DNA was
                     synthesized from mRNA using an XhoI-poly(dT)
                     linker-primer. EcoRI adopters were ligated to the
                     blunt-ended cDNA fragments and the products were digested
                     with XhoI for directional cloning into lambda ZAP Express
                     vector. This lambda library was amplified once using E.
                     coli host strain XL1 Blue MRF'. Inserts were then
                     subcloned by mass excision using ExAssist helper phage for
                     conversion into phagemid vector pBK-CMV in E. coli host
                     strain XLOLR."
BASE COUNT      279 a    150 c    200 g    206 t     10 others
ORIGIN

Query Match 60.6%; Score 700; DB 10; Length 845;
Best Local Similarity 92.8%; Pred. No. 1.1e-185;
Matches 772; Conservative 0; Mismatches 55; Indels 5; Gaps 4;

Qy    196 AGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAA 255
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      1 AGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCAAACAGAA 60

Qy    256 AATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGT 315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGT 120

Qy    316 TCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACA 375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    121 TCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACA 180

Qy    376 AGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGA 435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGA 240

Qy    436 AAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTT 495
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    241 AAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGGGCGTTTT 300

Qy    496 CTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTT 555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTT 360

Qy    556 CTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGT 615
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||| |
Db    361 CTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGATGGCACT 420

Qy    616 ATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTT 675
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    421 ATCAATCGTTCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATGATTATCCTT 480

Qy    676 ACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACG 735
          |||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db    481 ACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATNCAAGACAAGGGTACAATTGACC 540

Qy    736 GATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCG-TTC 794
          |||||  |||||||||||||||||||||||||| ||||||||||||||||||||| |||
Db    541 GATATTAAACTCTAATAATGTCAGATGAGAGTCCAGAATCAGAGACAGAGCAAGCGTTTC 600

Qy    795 TTAAGCGCCATCC-TTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTT 853
          ||||  ||||||  | ||||||||||| |||||||  ||  |||||||||||||| ||||||||
Db    601 TTAANCGCCATNCTTTGAGCAACCCATTAGTGNCTNCATTGATGCAAAAGAATTTCATTT 660

Qy    854 ATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTT 913
          ||  ||||||||||||||||||||| |||||||||  |||||||  |||  ||||||||||||
Db    661 ATTCACCGGGGGAATTTATGATGGGGAAAACTGTCCAAGTCCNTATNGGATTAATCACTT 720

Qy    914 TGTTTTACTTGTGGGTTATGGTTCAGC--GGATGGTGTAGATTACTGGATAG-CGAAAAA 970
          ||  ||  ||||||||  |||  |  |||  |   |  |||  |||  ||    ||||||  |||||||
Db    721 TGGTTACCTTGTGGGGTATTGGTCAACCGGATTGGNGTANATACTGGGATAGCCGAAAAA 780

Qy    971 TTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAAC 1022
          |||  |||||||||  |  |  |||    |  |   |  |  ||||||    |  ||  ||||
Db    781 TTCTTGGGGAGAAAAATTGGGGAGAAAATGGTCCATTTGATCCAAANAAAAC 832


RESULT 3
AI736102
LOCUS       AI736102                 756 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  sb22g10.y1 Gm-c1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1007-379 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AI736102
VERSION     AI736102.1  GI:5057626
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Possible reversed clone: similarity on wrong strand This clone is
            available through: ResGen, Invitrogen Corp. 2130 South Memorial
            Parkway Huntsville, AL 35801 For further information call: (800
            )-533-4363 or contact via email: ccu@resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .756
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1007-379"
                     /clone_lib="Gm-c1007"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from
                     immature cotyledons of greenhouse grown plants (individual
                     seed fresh weight of 100-300mg). The library was prepared
                     using the Life Technologies pSuperScript cDNA library
                     construction kit. Complementary DNA was synthesized from
                     mRNA using a poly (dT) sequence with a NotI restriction
                     site. SalI linkers adapters were ligated to the
                     blunt-ended cDNA fragments followed by NotI digestion. The
                     cDNA fragments were directionally cloned into the
                     NotI-SalI restriction site of the pSPORT1 vector. The
                     ligated cDNA fragments were transformed into E. coli
                     ElectroMax DH10B host cells. This library was constructed
                     by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      233 a    167 c    171 g    185 t
ORIGIN

  Query Match             59.9%;  Score 692;  DB 9;  Length 756;
  Best Local Similarity   97.2%;  Pred. No. 2e-183;
  Matches  704;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      2 CGGCCGCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTC 61
          |||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     33 CGGGGGTATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTC 92

Qy     62 CAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAA 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 CAGCATATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAA 152

Qy    122 ACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCA 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 ACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCA 212

Qy    182 CGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACAT 241
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||||
Db    213 CGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCACGGACAT 272

Qy    242 GAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACAT 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 GAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACAT 332

Qy    302 CACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAAT 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 CACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAAT 392

Qy    362 CAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATC 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATC 452

Qy    422 ATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAG 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAG 512

Qy    482 GGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGA 541
          |||||||||| |||||||||||||||||||||||||||||||||||||||||| ||||
```

```
Db    513 CGGTTGGGCGTGTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACATGAGA 572

Qy    542 CCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTA 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| ||
Db    573 CCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAATGTTCGTA 632

Qy    602 CAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGA 661
          |||||||||||| ||||||||| ||||||| ||| |||||||||||| | ||||||||||||
Db    633 CAATGGATGGCACTATCAATCGCTCGAATGAGTTCTAGAACATGGTCGGATTGCCACTGA 692

Qy    662 TGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAA 721
          |||||||||||||||||||||| |||||||||||||| ||||| | ||||||||||||||
Db    693 TGATGATTATCCTTACAGAGCTAGAGAGGGTAGATGCATAGCCATTCAGATACAAGACAA 752

Qy    722 GGTT 725
          ||||
Db    753 GGTT 756
```

RESULT 4
AI941099
LOCUS AI941099 729 bp mRNA linear EST 30-NOV-2001
DEFINITION sb84d09.y1 Gm-c1010 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1010-1674 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA sequence.
ACCESSION AI941099
VERSION AI941099.1 GI:5688084
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 729)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Possible reversed clone: similarity on wrong strand This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 382.

FEATURES Location/Qualifiers
source 1. .729
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1010-1674"
/clone_lib="Gm-c1010"
/tissue_type="young cotyledons of greenhouse grown plants"
/dev_stage="2cm long 12 week old"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from immature cotyledons (100-200mgs) of old greenhouse grown plants. The cDNA library was prepared using the Stratagene pBluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

BASE COUNT 228 a 158 c 169 g 173 t 1 others
ORIGIN

Query Match 59.7%; Score 690.2; DB 9; Length 729;
Best Local Similarity 97.4%; Pred. No. 6.3e-183;
Matches 701; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```
Qy     18 CTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATATCAACTCAT 77
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 CTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATATCAACTCAT 69

Qy     78 CGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACTG 137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 CGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACTG 129

Qy    138 TTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAAAG 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 TTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAAAG 189

Qy    198 AGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAAAA 257
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    190 AGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCAAACAGAAAA 249

Qy    258 TCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTC 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 TCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTC 309

Qy    318 AGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAG 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 AGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAG 369

Qy    378 AAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAA 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 AAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAA 429

Qy    438 AAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCT 497
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db    430 AAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGATGGGCGTTTTCT 489

Qy    498 GCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCT 557
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    490 GCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTATCT 549

Qy    558 GAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTAT 617
          ||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||| |||
Db    550 GAACATGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGATGGCACTAT 609

Qy    618 CAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTAC 677
          |||||||||| ||||||| |||||||||||| |||||||||||||||||||||||||| |||
Db    610 CAATCGTTCGGATGGGTTGTAGAACATGGTGGGATTGCCACTGATGATGATTATCCCTAC 669

Qy    678 AGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGA 737
          ||||||| ||| |||||||||| |||| ||| |||||| |||||||||||| | ||||||
Db    670 AGAGCTACAGACGGTAGATGCANAGCCCATATGATACATGACAAGGTTACACTCGACGGA 729
```

RESULT 5
AI941069
LOCUS AI941069 732 bp mRNA linear EST 30-NOV-2001
DEFINITION sb84a04.y1 Gm-c1010 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1010-1639 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA sequence.
ACCESSION AI941069
VERSION AI941069.1 GI:5688054
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Possible reversed clone: similarity on wrong strand This clone is
            available through: ResGen, Invitrogen Corp. 2130 South Memorial
            Parkway Huntsville, AL 35801 For further information call: (800
            )-533-4363 or contact via email: ccu@resgen.com
            High quality sequence stop: 391.
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1010-1639"
                     /clone_lib="Gm-c1010"
                     /tissue_type="young cotyledons of greenhouse grown plants"
                     /dev_stage="2cm long 12 week old"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature cotyledons (100-200mgs) of old greenhouse
                     grown plants. The cDNA library was prepared using the
                     Stratagene pBluescript II SK(+) library construction kit.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly (dT) sequence with a XhoI restriction
                     site. EcoRI adapters were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA fragments
                     were directionally cloned into the EcoRI-XhoI restriction
                     site of the pBluescript vector. The ligated cDNA fragments
                     were transformed into XL10-Gold host cells. This library
                     was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      228 a    164 c    167 g    173 t
ORIGIN

  Query Match             57.3%;  Score 662.4;  DB 9;  Length 732;
  Best Local Similarity   94.4%;  Pred. No. 4.1e-175;
  Matches  687;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       7 GCATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCA 66
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5 GTATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCA 64

Qy      67 TATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGG 126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65 TATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGG 124

Qy     127 TGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAG 186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 TGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAG 184

Qy     187 AAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATG 246
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     185 AAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATG 244

Qy     247 CAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTC 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 CAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTC 304

Qy     307 CTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAA 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 CTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAA 364

Qy     367 TGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGG 426
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 TGGCCAACAGGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGG 424

Qy     427 ATTGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTT 486
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     425 ATTGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTT 484

Qy     487 GGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTG 546
           | ||||| ||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     485 GAGCGTTCTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCATT 544

Qy     547 TTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATG 606
           ||||| ||||||||| |||||||||||||| | || |||||||| | ||| ||||||||
Db     545 GTAGCCATTCTGAACACGAACTCGTAGACTGCGCGGCAGAAAGCGGAAGTTGTTACAATG 604

Qy     607 GATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATG 666
           |||||| |||||||| ||| |  || |     |||||||| ||||||||||||||||
Db     605 CATGGCACTATCAATCGCTCGTAATGGGTCCTAGACATGGTGGTATTGCCACTGATGATG 664

Qy     667 ATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTA 726
           | ||| |||||||||||||| |||| | |||||| || | |||||||||||||| ||| ||
Db     665 ACTATGCTTACAGAGCTAAGGAGGTTGGATGCACAGTCCATAAGATACAAGACCAGGCTA 724

Qy     727 CAATTGAC 734
           ||| ||||
Db     725 CAACTGAC 732


RESULT 6
AI748608
LOCUS       AI748608                 705 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  sb59e03.y1 Gm-c1010 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1010-53 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AI748608
VERSION     AI748608.1  GI:5126872
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Possible reversed clone: similarity on wrong strand This clone is
            available through: ResGen, Invitrogen Corp. 2130 South Memorial
            Parkway Huntsville, AL 35801 For further information call: (800
            )-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1615   Std Error: 0.00
            High quality sequence stop: 392.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
```

```
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1010-53"
                     /clone_lib="Gm-c1010"
                     /tissue_type="young cotyledons of greenhouse grown plants"
                     /dev_stage="2cm long 12 week old"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature cotyledons (100-200mgs) of old greenhouse
                     grown plants. The cDNA library was prepared using the
                     Stratagene pBluescript II SK(+) library construction kit.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly (dT) sequence with a XhoI restriction
                     site. EcoRI adapters were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA fragments
                     were directionally cloned into the EcoRI-XhoI restriction
                     site of the pBluescript vector. The ligated cDNA fragments
                     were transformed into XL10-Gold host cells. This library
                     was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      220 a    150 c    156 g    177 t      2 others
ORIGIN

  Query Match             57.1%;  Score 659.8;  DB 9;  Length 705;
  Best Local Similarity   97.9%;  Pred. No. 2.2e-174;
  Matches  667;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      9 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 68
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 84

Qy     69 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 144

Qy    129 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 204

Qy    189 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 248
          |||||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||
Db    205 GAGGCAAAGAGACTTGAGATTTTCAAGATTAACTTGAACTATATCAGGGACATGAATGCA 264

Qy    249 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 324

Qy    309 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 384

Qy    369 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 444

Qy    429 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGG 488
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    445 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG 504

Qy    489 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 564

Qy    549 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA 608
          |||||||||||||||||||||||||||||||||||    | ||| ||||||||||||||
Db    565 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAAGCGAAGTTGTTACAATGGA 624

Qy    609 TGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT 668
          ||||| |||||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db    625 TGGCACTATCAATCGTTCGAATGGGTTNTAGAACATGGTGGGATTGCCACTGATGATGAT 684

Qy    669 TATCCTTACAGAGCTAAAGAG 689
          ||| |||||||||||| ||||
Db    685 TATNCTTACAGAGCTACAGAG 705
```

```
RESULT 7
AW472063
LOCUS       AW472063                704 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  si19e07.y1 Gm-c1029 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1029-1693 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AW472063
VERSION     AW472063.1  GI:7042169
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 704)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1508   Std Error: 0.00
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .704
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1029-1693"
                     /clone_lib="Gm-c1029"
                     /tissue_type="very young cotyledons of greenhouse grown
                     plants"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; This
                     cDNA library was constructed from mRNA isolated from very
                     young cotyledons (20-50mgs fresh weight) of greenhouse
                     grown plants.  The library was prepared using the Life
                     Technologies pSuperScript cDNA library construction kit.
                     Complementary DNA was synthesized from mRNA using a
                     poly(dT) sequence with a Not I restriction site.  Sal I
                     linker adapters were ligated to the blunt-ended cDNA
                     fragments followed by Not I digestion.  The cDNA fragments
                     were directionally cloned into the Not I-Sal I restriction
                     site of the pSPORT1 vector.  The ligated cDNA fragments
                     were transformed into E.coli Electromax DH10B host cells.
                     This library was constructed by Dr. Paul Keim and Dr.
                     Virginia Coryell."
BASE COUNT      220 a    148 c    155 g    180 t      1 others
ORIGIN

  Query Match             56.7%;  Score 655.8;  DB 10;  Length 704;
  Best Local Similarity   97.4%;  Pred. No. 2.9e-173;
  Matches  666;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      9 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 68
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 80

Qy     69 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      81 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 140

Qy     129 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 200

Qy     189 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 248
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     201 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCA 260

Qy     249 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 320

Qy     309 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 380

Qy     369 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 440

Qy     429 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGG 488
           |||||||||||||||||||||||||||||||| |||||||||||| ||||||
Db     441 TGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAATGGGGCTGTGGAAGCGGTTGG 500

Qy     489 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 548
           |||| ||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     501 GCGTATTCTGCCACGGGAGCCATAGAAGCAGCACATGCTATAGCAACAGGAGACCTTGTT 560

Qy     549 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA 608
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     561 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGA 620

Qy     609 TGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT 668
           ||||| |||||||| |||||||| ||| |||||||||||||| || ||||||||||||||
Db     621 TGGCACTATCAATCAGTCGAATGTGTTCTAGAACATGGTGGGATGGCCACTGATGATGAA 680

Qy     669 TATCCTTACAGAGCTAAAGAGGGT 692
           ||| ||||| || ||||||| |||
Db     681 TATTCTTACTGANCTAAAGATGGT 704


RESULT 8
BE660304
LOCUS       BE660304                664 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  2-C10 GmaxSC Glycine max cDNA, mRNA sequence.
ACCESSION   BE660304
VERSION     BE660304.1  GI:9986196
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Harris,N., Chapman,B.P. and Gijzen,M.
  TITLE     Gene expression in developing soybean seed coats
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Gijzen M
            Agriculture and Agri-Food Canada
            1391 Sandford Street, London, Ontario, Canada N5V 4T3
            Tel: 519 457 1470
            Fax: 519 457 3997
            Email: gijzenm@em.agr.ca.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Glycine max"
                     /cultivar="Harosoy 63"
                     /db_xref="taxon:3847"
                     /clone_lib="GmaxSC"
                     /tissue_type="Seed coats"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
                     cDNA library was constructed from polyA+ enriched mRNA
                     from  green seed coats in mid to late developemental stage
                     , average fresh  weight 250 mg per seed. Traces of pod and
                     embryo tissue also present.  Complementary DNA was
                     synthesized from mRNA using an XhoI-poly(dT)
                     linker-primer. EcoRI adopters were ligated to the
                     blunt-ended cDNA  fragments and the products were digested
                     with XhoI for directional  cloning into lambda ZAP Express
                     vector. This lambda library was  amplified once using E.
                     coli host strain XL1 Blue MRF'. Inserts were  then
                     subcloned by mass excision using ExAssist helper phage for
                     conversion into phagemid vector pBK-CMV in E. coli host
                     strain XLOLR."
BASE COUNT      223 a    132 c    155 g    151 t      3 others
ORIGIN

  Query Match           56.2%;  Score 649.8;  DB 10;  Length 664;
  Best Local Similarity   98.5%;  Pred. No. 1.4e-171;
  Matches  654;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      67 TATCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGG 126
           |||||||||||||||||||||||  |||||||||||||||||||||||||||||| ||||
Db       1 TATCAACTCATCGTTCCATATTGGNTCTTGACCTAACCAAGTTTACCACACAGAANCAGG 60

Qy     127 TGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAG 186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAG 120

Qy     187 AAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATG 246
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     121 AAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATG 180

Qy     247 CAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTC 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTC 240

Qy     307 CTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAA 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAA 300

Qy     367 TGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGG 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGG 360

Qy     427 ATTGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTT 486
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     361 ATTGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTT 420

Qy     487 GGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTG 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTG 480

Qy     547 TTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATG 606
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     481 TTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATG 540

Qy     607 GATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATG 666
           ||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     541 GATGGCACTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATG 600

Qy     667 ATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTA 726
           |||||||||||||||||||||||||||||||||||||| |||||||||||||| ||
Db     601 ATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATANGATACAAGACAAGGGTA 660

Qy     727 CAAT 730
           ||||
Db     661 CAAT 664


RESULT 9
AW317682
```

```
LOCUS       AW317682                659 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sg56a06.y1 Gm-c1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1007-395 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AW317682
VERSION     AW317682.1  GI:6747226
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1007-395"
                     /clone_lib="Gm-c1007"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from
                     immature cotyledons of greenhouse grown plants (individual
                     seed fresh weight of 100-300mg). The library was prepared
                     using the Life Technologies pSuperScript cDNA library
                     construction kit. Complementary DNA was synthesized from
                     mRNA using a poly (dT) sequence with a NotI restriction
                     site. SalI linkers adapters were ligated to the
                     blunt-ended cDNA fragments followed by NotI digestion. The
                     cDNA fragments were directionally cloned into the
                     NotI-SalI restriction site of the pSPORT1 vector. The
                     ligated cDNA fragments were transformed into E. coli
                     ElectroMax DH10B host cells. This library was constructed
                     by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      218 a    124 c    157 g    159 t      1 others
ORIGIN

  Query Match             56.1%;  Score 648.4;  DB 10;  Length 659;
  Best Local Similarity   98.9%;  Pred. No. 3.5e-171;
  Matches  652;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    262 CCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCA 321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCA 60

Qy    322 AAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAA 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAA 120

Qy    382 TGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAG 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAG 180

Qy    442 GTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCA 501
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    181 GTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGGGCGTTTTCTGCCA 240

Qy    502 CGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAAC 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAAC 300

Qy    562 AAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAAT 621
          |||||||||||||||||||||||||||||||||||| ||||||||||||||| |||||||
Db    301 AAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGATGGCACTATCAAT 360

Qy    622 CGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAG 681
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    361 CGTTCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATGATTATCCTTACAGAG 420

Qy    682 CTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATG 741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATG 480

Qy    742 AAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCG 801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCG 540

Qy    802 CCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCG 861
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    541 CCATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAGATTTTCATTTATACACCC 600

Qy    862 GGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTA 920
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    601 GGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGNTTTA 659


RESULT 10
AW317681
LOCUS       AW317681                657 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sg56a05.y1 Gm-c1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1007-393 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AW317681
VERSION     AW317681.1  GI:6747225
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Glycine max"
```

```
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1007-393"
                     /clone_lib="Gm-c1007"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from
                     immature cotyledons of greenhouse grown plants (individual
                     seed fresh weight of 100-300mg). The library was prepared
                     using the Life Technologies pSuperScript cDNA library
                     construction kit. Complementary DNA was synthesized from
                     mRNA using a poly (dT) sequence with a NotI restriction
                     site. SalI linkers adapters were ligated to the
                     blunt-ended cDNA fragments followed by NotI digestion. The
                     cDNA fragments were directionally cloned into the
                     NotI-SalI restriction site of the pSPORT1 vector. The
                     ligated cDNA fragments were transformed into E. coli
                     ElectroMax DH10B host cells. This library was constructed
                     by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      229 a    123 c    158 g    144 t      3 others
ORIGIN

  Query Match             55.9%;  Score 646;  DB 10;  Length 657;
  Best Local Similarity   98.8%;  Pred. No. 1.6e-170;
  Matches  649;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy     129 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 60

Qy     189 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 248
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      61 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCA 120

Qy     249 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 180

Qy     309 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 240

Qy     369 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 300

Qy     429 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGG 488
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     301 TGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG 360

Qy     489 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 420

Qy     549 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA 608
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     421 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGA 480

Qy     609 TGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT 668
           ||||| |||||||||||||||||  ||||||||||  ||||||||||||||||||||||
Db     481 TGGCACTATCAATCGTTCGAATGGGTNNTAGAACATGGTGNGATTGCCACTGATGATGAT 540

Qy     669 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA 728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA 600

Qy     729 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAG 785
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAG 657


RESULT 11
AW433000
LOCUS       AW433000                655 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  si03b11.y1 Gm-c1029 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1029-118 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AW433000
VERSION     AW433000.1  GI:6964307
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1190   Std Error: 0.00
            High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1029-118"
                     /clone_lib="Gm-c1029"
                     /tissue_type="very young cotyledons of greenhouse grown
                     plants"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; This
                     cDNA library was constructed from mRNA isolated from very
                     young cotyledons (20-50mgs fresh weight) of greenhouse
                     grown plants.  The library was prepared using the Life
                     Technologies pSuperScript cDNA library construction kit.
                     Complementary DNA was synthesized from mRNA using a
                     poly(dT) sequence with a Not I restriction site.  Sal I
                     linker adapters were ligated to the blunt-ended cDNA
                     fragments followed by Not I digestion.  The cDNA fragments
                     were directionally cloned into the Not I-Sal I restriction
                     site of the pSPORT1 vector.  The ligated cDNA fragments
                     were transformed into E.coli Electromax DH10B host cells.
                     This library was constructed by Dr. Paul Keim and Dr.
                     Virginia Coryell."
BASE COUNT      226 a    122 c    156 g    150 t      1 others
ORIGIN

  Query Match             55.0%;  Score 635.4;  DB 10;  Length 655;
  Best Local Similarity   98.2%;  Pred. No. 1.6e-167;
  Matches  642;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     194 AAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAG 253
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1 AAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCAAACAG 60

Qy     254 AAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGA 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGA 120

Qy     314 GTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAA 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    121 GTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAA 180

Qy    374 CAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAG 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAG 240

Qy    434 GAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTT 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    241 GAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGGGCGTT 300

Qy    494 TTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCT 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCT 360

Qy    554 TTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCA 613
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    361 TTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGATGGCA 420

Qy    614 GTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCC 673
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    421 CTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATGATTATCC 480

Qy    674 TTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGA 733
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGA 540

Qy    734 CGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTT 793
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTT 600

Qy    794 CTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTT 847
          |||||||||||| ||||||| ||||| |||||||| |||||||| |||||||
Db    601 CTTAAGCGCCATNCTTGAGCGACCAAGTAGTGTCTACATTGATGCAGAAGATTT 654


RESULT 12
AW398014
LOCUS       AW398014                657 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sg71b10.y1 Gm-c1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1007-1844 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AW398014
VERSION     AW398014.1  GI:6916484
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1074   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 437.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1007-1844"
                     /clone_lib="Gm-c1007"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from
                     immature cotyledons of greenhouse grown plants (individual
                     seed fresh weight of 100-300mg). The library was prepared
                     using the Life Technologies pSuperScript cDNA library
                     construction kit. Complementary DNA was synthesized from
                     mRNA using a poly (dT) sequence with a NotI restriction
                     site. SalI linkers adapters were ligated to the
                     blunt-ended cDNA fragments followed by NotI digestion. The
                     cDNA fragments were directionally cloned into the
                     NotI-SalI restriction site of the pSPORT1 vector. The
                     ligated cDNA fragments were transformed into E. coli
                     ElectroMax DH10B host cells. This library was constructed
                     by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      214 a    113 c    170 g    159 t      1 others
ORIGIN

  Query Match             54.8%;  Score 633.6;  DB 10;  Length 657;
  Best Local Similarity   97.7%;  Pred. No. 5e-167;
  Matches  642;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy    327 TACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAG 386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAG 60

Qy    387 AAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTC 446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTC 120

Qy    447 ATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGA 506
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    121 ATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGGGCGTTTTCTGCCACGGGA 180

Qy    507 GCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAA 566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAA 240

Qy    567 CTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTC 626
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db    241 CTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGATGGCACTATCAATCGTTC 300

Qy    627 GAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAA 686
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    301 GAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATGATTATCCTTACAGAGCTAAA 360

Qy    687 GAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACT 746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACT 420

Qy    747 CTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATC 806
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATC 480

Qy    807 CTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGA 866
          |||||||||||||||||||  ||||||||| ||||| |||||||||||||| |||||
Db    481 CTTGAGCAACCAATTAGTGTCTGCATTGATGCANAAGATGTTCATTTATACACCCGGGGA 540

Qy    867 ATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTG 926
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    541 ATTTATGATGGAGAAAACTGTACAAGTGCGTATGGGATTAATCACTTTGTTTTACTTGTG 600

Qy    927 GGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAA 983
          |||||||||||||||||||||||| |||  |||||||| ||||||| |||||||
Db    601 GGTTATGGTTCAGCGGATGGTGTAGAATACGTGATAGCGAACAATTCATTGGGAGAA 657


RESULT 13
```

```
AI794978
LOCUS       AI794978                665 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  sb74e04.y1 Gm-c1010 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1010-727 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AI794978
VERSION     AI794978.1  GI:5342694
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 665)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Possible reversed clone: similarity on wrong strand This clone is
            available through: ResGen, Invitrogen Corp. 2130 South Memorial
            Parkway Huntsville, AL 35801 For further information call: (800
            )-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1178   Std Error: 0.00
            High quality sequence stop: 400.
FEATURES             Location/Qualifiers
     source          1. .665
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1010-727"
                     /clone_lib="Gm-c1010"
                     /tissue_type="young cotyledons of greenhouse grown plants"
                     /dev_stage="2cm long 12 week old"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature cotyledons (100-200mgs) of old greenhouse
                     grown plants. The cDNA library was prepared using the
                     Stratagene pBluescript II SK(+) library construction kit.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly (dT) sequence with a XhoI restriction
                     site. EcoRI adapters were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA fragments
                     were directionally cloned into the EcoRI-XhoI restriction
                     site of the pBluescript vector. The ligated cDNA fragments
                     were transformed into XL10-Gold host cells. This library
                     was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      214 a    130 c    157 g    161 t      3 others
ORIGIN

  Query Match             54.5%;  Score 630;  DB 9;  Length 665;
  Best Local Similarity   96.5%;  Pred. No. 5.2e-166;
  Matches  642;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     265 ATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAA 324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAA 60

Qy     325 AGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGA 384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGA 120

Qy     385 AGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTG 444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     121 AGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAAGTG 180

Qy     445 TCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGG 504
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     181 TCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGGGCGTTTTCTGCCACGG 240

Qy     505 GAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAG 564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAG 300

Qy     565 AACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGT 624
           |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||||
Db     301 AACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTTACAATGGATGGCACTATCAATCGT 360

Qy     625 TCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTA 684
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     361 TCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATGATTATCCTTACAGAGCTA 420

Qy     685 AAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAA 744
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     421 AAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGATA 480

Qy     745 CTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCA 804
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTCTTATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCA 540

Qy     805 TCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGG 864
           | |||||||||||||| ||||||||| |||||||||||||||| |||||||||||  | | |
Db     541 TNCTTGAGCAACCAACTAGTGTCTCCATTGATGCAAAAGATCTTCATTTATACCACCGNG 600

Qy     865 GAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTG 924
           | || |||||||||||||||||| | ||||||||| |||||||||||| |||| ||||||
Db     601 GCATCTATGATGGAGAAAACTGTCCCAGTCCGTATCGGATTAATCACTCTGTTCTACTTG 660

Qy     925 TGGGT 929
           | |||
Db     661 TNGGT 665


RESULT 14
AW397984
LOCUS       AW397984                654 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sg70g06.y1 Gm-c1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1007-1811 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AW397984
VERSION     AW397984.1  GI:6916454
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
```

```
          South Memorial Parkway Huntsville, AL 35801 For further information
          call: (800)-533-4363 or contact via email: ccu@resgen.com
          Seq primer: -40RP from Gibco
          High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .654
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1007-1811"
                     /clone_lib="Gm-c1007"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from
                     immature cotyledons of greenhouse grown plants (individual
                     seed fresh weight of 100-300mg). The library was prepared
                     using the Life Technologies pSuperScript cDNA library
                     construction kit. Complementary DNA was synthesized from
                     mRNA using a poly (dT) sequence with a NotI restriction
                     site. SalI linkers adapters were ligated to the
                     blunt-ended cDNA fragments followed by NotI digestion. The
                     cDNA fragments were directionally cloned into the
                     NotI-SalI restriction site of the pSPORT1 vector. The
                     ligated cDNA fragments were transformed into E. coli
                     ElectroMax DH10B host cells. This library was constructed
                     by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      208 a    139 c    148 g    158 t      1 others
ORIGIN

  Query Match             54.2%;  Score 626.4;  DB 10;  Length 654;
  Best Local Similarity   97.4%;  Pred. No. 5.3e-165;
  Matches  636;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      33 TTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATATCAACTCATCGTTCCATATTGGAC 92
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATATCAACTCATCGTTCCATATTGGAC 60

Qy      93 CTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAG 152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAG 120

Qy     153 AGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTC 212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTC 180

Qy     213 AAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCAT 272
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGAATAACTTGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCAT 240

Qy     273 CGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTG 332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTG 300

Qy     333 CAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAA 392
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAA 360

Qy     393 CAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACC 452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACC 420

Qy     453 CAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATA 512
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     421 CAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGGGCGTTTTCTGCCACGGGAGCCATA 480

Qy     513 GAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTA 572
           |||||||||||||||||||||||||||||||  |||||| ||| ||||||||||||||
Db     481 GAAGCAGCACATGCAATAGCAACAGGAGACCTTAGTAGCCTATCTTGACAAGAACTCGTA 540

Qy     573 GACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGG 632
           |||||||||||||||||| || | ||||||||||||||| ||||||||||||||||||||
Db     541 GACTGTGTGGAAGAAAGCGANGGGTGTTACAATGGATGGCACTATCAATCGTTCGAATGG 600

Qy     633 GTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAA 685
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTTTTTAGACATGGTGGGATTGCCACTGATGATGATTATCCTTACAGACTTAA 653


RESULT 15
AW318029
LOCUS       AW318029                677 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sg60c09.y1 Gm-c1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1007-809 5' similar to TR:O64458 O64458 BD 30K. ;, mRNA
            sequence.
ACCESSION   AW318029
VERSION     AW318029.1  GI:6747573
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1043   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1007-809"
                     /clone_lib="Gm-c1007"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from
                     immature cotyledons of greenhouse grown plants (individual
                     seed fresh weight of 100-300mg). The library was prepared
                     using the Life Technologies pSuperScript cDNA library
                     construction kit. Complementary DNA was synthesized from
                     mRNA using a poly (dT) sequence with a NotI restriction
                     site. SalI linkers adapters were ligated to the
                     blunt-ended cDNA fragments followed by NotI digestion. The
                     cDNA fragments were directionally cloned into the
                     NotI-SalI restriction site of the pSPORT1 vector. The
                     ligated cDNA fragments were transformed into E. coli
                     ElectroMax DH10B host cells. This library was constructed
                     by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      212 a    147 c    148 g    167 t      3 others
ORIGIN

  Query Match             54.1%;  Score 625;  DB 10;  Length 677;
  Best Local Similarity   96.5%;  Pred. No. 1.3e-164;
  Matches  637;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       9 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 68
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18 ATGGGTTTCCTTGTGTTGCTTCTTTTCTCCCTCTTAGGTCTCTCTTCTAGTTCCAGCATA 77
```

```
Qy      69 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG 137

Qy     129 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 TCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAA 197

Qy     189 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCA 248
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     198 GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGAACTATATCAGGGACATGAATGCA 257

Qy     249 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT 317

Qy     309 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     318 CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATG 377

Qy     369 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     378 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 437

Qy     429 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGG 488
           |||||||||||||||||||||||||||||||||||||||||||||||||||| || |||
Db     438 TGGAGGAAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGATGG 497

Qy     489 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 548
           |||| ||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     498 GCGTNTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAAGAGACCTTGTT 557

Qy     549 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA 608
           |||||||||||||||||||||||||||||||||||||| |||||| ||| ||||||||||
Db     558 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGATAGCGAACGTTAGTACAATGGA 617

Qy     609 TGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT 668
           ||||| ||||||    |||||  ||||||||||    |||  || ||||||||||||||
Db     618 TGGCACTATCAAATCGTCGAAATGGTTTTAGAAACATGTGNGATNGCCACTGATGATGAT 677
```

Search completed: January 11, 2003, 21:09:52
Job time : 1933 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 19:47:27 ; Search time 55 Seconds
(without alignments)
6445.788 Million cell updates/sec

Title: US-09-805-694B-1
Perfect score: 1156
Sequence: 1 gcggccgcatgggtttcctt.....tctcctctttgagcggccgc 1156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	154.4	13.4	1056	1	US-09-500-651-1	Sequence 1, Appli
2	154.4	13.4	1056	1	US-08-813-591-1	Sequence 1, Appli
3	151.2	13.1	1577	4	US-08-821-994-59	Sequence 59, Appl
4	147	12.7	1553	4	US-08-821-994-60	Sequence 60, Appl
5	134.2	11.6	1062	4	US-09-325-932A-109	Sequence 109, App
6	132.8	11.5	1600	4	US-09-325-932A-98	Sequence 98, Appl
7	128.6	11.1	2167	4	US-09-325-932A-102	Sequence 102, App
8	120.2	10.4	1434	4	US-08-821-994-62	Sequence 62, Appl
9	118.6	10.3	1390	4	US-08-821-994-61	Sequence 61, Appl
10	118.6	10.3	1441	4	US-08-821-994-63	Sequence 63, Appl
11	117.2	10.1	765	4	US-09-325-932A-110	Sequence 110, App
12	117	10.1	1102	4	US-08-821-994-86	Sequence 86, Appl
13	117	10.1	1186	1	US-08-453-924-9	Sequence 9, Appli
14	115.2	10.0	1661	4	US-08-821-994-82	Sequence 82, Appl
15	114.8	9.9	1474	4	US-08-821-994-64	Sequence 64, Appl
16	98.8	8.5	1474	4	US-09-325-932A-101	Sequence 101, App
17	97.6	8.4	883	4	US-09-325-932A-96	Sequence 96, Appl
18	97.6	8.4	948	4	US-09-325-932A-95	Sequence 95, Appl
19	92.6	8.0	1407	4	US-09-325-932A-107	Sequence 107, App
20	92.6	8.0	1463	4	US-09-325-932A-92	Sequence 92, Appl
21	91.8	7.9	1334	2	US-08-560-398-9	Sequence 9, Appli
22	91.6	7.9	388	4	US-09-325-932A-108	Sequence 108, App
23	91	7.9	441	4	US-09-325-932A-122	Sequence 122, App
24	87.4	7.6	1066	4	US-09-325-932A-97	Sequence 97, Appl
25	82.8	7.2	392	4	US-09-325-932A-113	Sequence 113, App
26	80	6.9	1301	2	US-08-641-314C-1	Sequence 1, Appli
27	77.2	6.7	811	4	US-09-004-731-88	Sequence 88, Appl
28	77.2	6.7	811	4	US-08-749-699-88	Sequence 88, Appl
29	77.2	6.7	811	4	US-09-004-729-88	Sequence 88, Appl
c 30	76.4	6.6	823	4	US-09-004-731-90	Sequence 90, Appl
c 31	76.4	6.6	823	4	US-08-749-699-90	Sequence 90, Appl
c 32	76.4	6.6	823	4	US-09-004-729-90	Sequence 90, Appl
33	73.4	6.3	5185	4	US-08-971-395-5	Sequence 5, Appli
34	72.2	6.2	1643	4	US-09-701-685-1	Sequence 1, Appli
35	72	6.2	762	4	US-09-004-731-91	Sequence 91, Appl
36	72	6.2	762	4	US-08-749-699-91	Sequence 91, Appl
37	72	6.2	762	4	US-09-004-729-91	Sequence 91, Appl
38	71.4	6.2	475	4	US-09-325-932A-120	Sequence 120, App
c 39	71.4	6.2	783	4	US-09-004-731-93	Sequence 93, Appl
c 40	71.4	6.2	783	4	US-08-749-699-93	Sequence 93, Appl
c 41	71.4	6.2	783	4	US-09-004-729-93	Sequence 93, Appl
42	71.2	6.2	660	4	US-09-004-731-94	Sequence 94, Appl
43	71.2	6.2	660	4	US-08-749-699-94	Sequence 94, Appl
44	71.2	6.2	660	4	US-09-004-729-94	Sequence 94, Appl
45	70.8	6.1	700	4	US-09-325-932A-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1
US-09-500-651-1
; Sequence 1, Application US/09500651
; GENERAL INFORMATION:
; APPLICANT: ASANO, MINAO
; APPLICANT: KAWAI, MISAKO
; APPLICANT: MIWA, TETSUYA
; APPLICANT: NIO, NORIKI
; TITLE OF INVENTION: AMINOPEPTIDASE GX, AND A METHOD OF
; TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,651
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/813,591
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 030458/1997
; FILING DATE: 14-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-845-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Glycine max

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1056
;      OTHER INFORMATION:  /note= "METHOD OF DETERMINING THE
;      OTHER INFORMATION:  CHARACTERISTICS: P"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..9
;      OTHER INFORMATION:  /note= "INSERTION SEQUENCE, METHOD
;      OTHER INFORMATION:  OF DETERMINING THE CHARACTERISTICS: E"
US-09-500-651-1

  Query Match            13.4%;  Score 154.4;  DB 1;  Length 1056;
  Best Local Similarity  52.9%;  Pred. No. 1e-38;
  Matches  494;  Conservative    0;  Mismatches  401;  Indels   39;  Gaps    6;

Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
          ||| |||   |||  || |  || ||   ||  ||| | | |  |||       |||| |
Db     38 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 97

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
          |    || ||| |    | ||  ||  |||  ||| |  |  |||| |||||| | |||
Db     98 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 157

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
          |   |  |||   |||   ||  |         |   | || |    |      | |||
Db    158 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 211

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
          |||||  ||| |||||  ||||     | || | |||    |||||||||  | |    |
Db    212 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 271

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
            |||       | |    ||  || | ||   |  ||    |  |  ||   |
Db    272 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 328

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
          | ||| |    ||||  ||    ||||||||||  ||||||   |  | |  || ||
Db    329 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 388

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
          ||||||| |||||||| ||  |||||||| || || || |  || ||  |||||| |  |
Db    389 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 448

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
          ||   |||| |||||| || ||  ||    | || |||||||| | || || ||||
Db    449 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 508

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
            | | | | |||| |   |    |||      | |  |  | || || |   |  |
Db    509 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 568

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
          |  |||| |||||||   |||||||| ||||| || ||| | | |   || |||||||||
Db    569 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 628

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756
          |  |  ||| || | | |  ||    ||| | ||||| |  || |||  | |       |
Db    629 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 688

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
            ||||||                        ||| ||   ||  ||| |     | || |
Db    689 ATGATGAGTT--------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 727

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
          |  | || ||  | |||||  | |   |       | ||  |       |   | | || ||
Db    728 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 784

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
            |   |   | | | ||      |   |  ||||   |||  |   ||| || |||||
Db    785 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 844

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
          |||| | |||   ||||| ||||| |  |  ||||||||||       |||||||| |
Db    845 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 904

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
          |||| |||||  |  |  |||||||    | |||
Db    905 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 938


RESULT 2
US-08-813-591-1
; Sequence 1, Application US/08813591
; Patent No. 5824534
;  GENERAL INFORMATION:
;    APPLICANT:  ASANO, MINAO
;    APPLICANT:  KAWAI, MISAKO
;    APPLICANT:  MIWA, TETSUYA
;    APPLICANT:  NIO, NORIKI
;    TITLE OF INVENTION:  AMINOPEPTIDASE GX, AND A METHOD OF
;    TITLE OF INVENTION:  HYDROLYSING A PROTEIN WITH THE SAME
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/813,591
;      FILING DATE:  07-MAR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 051848/1996
;      FILING DATE:  08-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 030458/1997
;      FILING DATE:  14-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-845-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1056 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Glycine max
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1056
;      OTHER INFORMATION:  /note= "METHOD OF DETERMINING THE
;      OTHER INFORMATION:  CHARACTERISTICS: P"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..9
;      OTHER INFORMATION:  /note= "INSERTION SEQUENCE, METHOD
;      OTHER INFORMATION:  OF DETERMINING THE CHARACTERISTICS: E"
US-08-813-591-1

  Query Match            13.4%;  Score 154.4;  DB 1;  Length 1056;
  Best Local Similarity  52.9%;  Pred. No. 1e-38;
```

Matches 494; Conservative 0; Mismatches 401; Indels 39; Gaps 6;

```
Qy    103 CCAAGTTTACCACACAGAAACAGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATG 162
Db     38 CCACGTTGCGCACCGAGGAGGAGCTGATGTCCATGTACGAGCAGTGGCTCGTGAAGCACG 97

Qy    163 GACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACT 222
Db     98 GGAAGGTGTACAACGCGCTCGGCGAGAAGGAGAAGCGCTTCCAGATCTTCAAGGACAACC 157

Qy    223 CGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGAT 282
Db    158 TGCGATTCATCGACGACCACAACTCCGCGGAGGACCGAACCTACAAGC------TCGGAC 211

Qy    283 TGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCA 342
Db    212 TGAACCGGTTCGCTGATCTCACCAACGAGGAATACAGGGCCAAGTACTTGGGAACCAAGA 271

Qy    343 AGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTT 402
Db    272 TCGATCCCAACCGGAGGCTCGGAAAGACCCCGAGCAACCGCTACGCGCCACG---TGTCG 328

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
Db    329 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 388

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
Db    389 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 448

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
Db    449 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 508

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
Db    509 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 568

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
Db    569 ACAATGGCGGCATTGATTCTGATGAGGATTACCCATACCGTGGTGTTGATGGTAGATGCG 628

Qy    700 AAGCCAATAAGATACAAGACAAG---GTTACAATTGACGGATATGAAACTCTAATAATGT 756
Db    629 ACACATATAGGAAAAATGCTAAAGTCGTTTCTATTGATGACTACGAAGATGTTCCTGCCT 688

Qy    757 CAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAAC 816
Db    689 ATGATGAGTT----------------------AGCCTTGAAAAAGGCCGTTGCAAATCAGC 727

Qy    817 CAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATG 876
Db    728 CCGTGAGCGTTGCTATTGAAGGAGGGGGCAGGGAATTTCAATTATATG---TATCTGGTG 784

Qy    877 GAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTT 936
Db    785 TATTCACGGGGAGATGTGGCACAGCACTAGATCATGGTGTCGTGGCTGTTGGGTATGGAA 844

Qy    937 CAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAG 996
Db    845 CAGCTAAAGGTCATGATTATTGGATCGTAAGGAATTCATGGGGTTCTAGCTGGGGAGAGG 904

Qy    997 ATGGTTACATTTGGATCCAAAGAAACACGGGTAA 1030
Db    905 ATGGCTACATCAGATTAGAAAGAAATCTTGCTAA 938
```

RESULT 3
US-08-821-994-59
; Sequence 59, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 1577
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-59

Query Match 13.1%; Score 151.2; DB 4; Length 1577;
Best Local Similarity 55.2%; Pred. No. 1.3e-37;
Matches 369; Conservative 1; Mismatches 274; Indels 24; Gaps 3;

```
Qy    414 CCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGC 473
Db    402 CCAGATTCTGTTGATTGGAGGAAGAAAGGAGCTGTTACTAATGTCAAAGATCAAGGAAGC 461

Qy    474 TGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCA 533
Db    462 TGCGGAGCGTGTTGGTCTTTCTCGGCGACTGGAGCTATGGAAGGAATCAACCAGATTGTA 521

Qy    534 ACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAA 593
Db    522 ACAGGAGATCTCATCAGCCTCTCTGAGCAAGAACTCATTGATTGTGATAAGTCATACAAC 581

Qy    594 GGTTCTTACAATGGATGGC---AGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGC 650
Db    582 GATGGATGCAATGGTGGTCTCATGGACTACGCTTTTCAATTTGTCATTAAAAACCATGGG 641

Qy    651 ATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAG 710
Db    642 ATTGACACAGAGAAAGATTATCCTTATCAAGAACGTGATGGCACCTGTMAGAAAGATAAG 701

Qy    711 ATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACA 770
Db    702 TTGAATAGAAAGGTTGTGACAATTGATAGCTACGCTGGTGTAAAATCAAATGACGAGAAA 761

Qy    771 GAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCA 830
Db    762 GCGTTACTAGAAGCTGTAGCGGCTCAGCCAGTTAGTGTTG-----------GTATCTGT 809

Qy    831 ATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACA 890
Db    810 GGGAGCGAGAGAGCGTTTCAGTTATACTCTAAGGGAATATTCTCTGGCCCATGTTCAACA 869

Qy    891 AGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTA 950
Db    870 TCATTG---------GACCACGCAGTGCTCATCGTAGGATACGGTTCAAAGAACGGTGTT 920

Qy    951 GATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGG 1010
Db    921 GATTACTGGATCGTGAAGAACTCTTGGGGAAAGAGTTGGGGAATGGATGGGTTTATCCAC 980

Qy   1011 ATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATAC 1070
Db    981 ATGCAGCGTAACACCGGCAACGCAGAAGGAGTATGCGGAATCAACATGCTGGCTTCATAT 1040

Qy   1071 CCAACCAA 1078
Db   1041 CCCATCAA 1048
```

RESULT 4
US-08-821-994-60

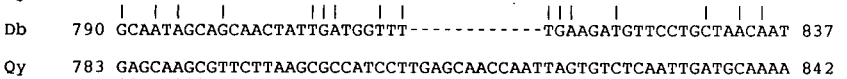
```
; Sequence 60, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
;   LENGTH: 1553
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-60

  Query Match             12.7%;  Score 147;  DB 4;  Length 1553;
  Best Local Similarity   54.6%;  Pred. No. 2.7e-36;
  Matches  372;  Conservative    0;  Mismatches  285;  Indels   24;  Gaps    3;

Qy    401 TTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAA 460
          | | | | |    || |  ||    |||||||||||| ||||| |   | ||  | || ||
Db    397 TGGGGGCAAAGTTCCAGATTCTGTTGATTGGAGGAAGAAAGGAGCTGTTACTAATGTCAA 456

Qy    461 GTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGC 520
          | |||||  |||| |||  | ||||| | || || || || ||||| || |||| |
Db    457 AGATCAAGGAAGCTGCGGAGCGTGTTGGTCTTTCTCGGCGACTGGAGCAATGGAAGGAAT 516

Qy    521 ACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGT 580
          |    || | ||||||||| ||  | ||||| ||||| || |||||  | || ||||
Db    517 CAACCAGATTGTAACAGGAGATCTCATCAGCCTCTCTGAGCAGGAACTAATTGATTGTGA 576

Qy    581 GGAAGAAAGCGAAGGTTCTTACAATGGATGGC---AGTATCAATCGTTCGAATGGGTTTT 637
          |      | | | |   | ||||||  | |    | |  |  | || |||  ||  |
Db    577 CAAGTCCTACAACGATGGATGCAATGGTGGTCTCATGGACTACGCTTTTCAATTTGTCAT 636

Qy    638 AGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATG 697
          || |   ||| || | ||| ||  | ||||||||||   ||     || || |  ||
Db    637 TAAAAACCATGGAATCGACACAGAGAAAGATTATCCTTATCAAGAACGTGATGGCACCTG 696

Qy    698 CAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTC 757
          |||    ||||| |  ||   ||||||   |      |      |  ||    |||  ||
Db    697 TAAAAAAGATAAGTTGAAAAGAAAGGTTGTGACAATTGATAGCTATGCTGGCGTAAAATC 756

Qy    758 AGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACC 817
          | | ||    | ||  | |  |  ||    |||          |  |   |||
Db    757 AAACGACGAGAAAGCGTTACTGGAAGCTGTAGCGGCTCAGCCAGTTAGTGTTG------- 809

Qy    818 AATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGG 877
              | |||         | | ||  ||||| ||||| |   ||||||  |    ||
Db    810 -----GCATCTGTGGCAGCGAGAGAGCGTTTCAGCTATACTCTAAGGGAATATTCTCTG- 863

Qy    878 AGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTC 937
              |  || || |        |  | |||   ||  |  | || || || ||||| 
Db    864 --------GCCCATGTTCAACATCATTGGACCACGCAGTGCTCATCGTAGGATACGGTTC 915

Qy    938 AGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA 997
          |  | | ||||| |||||||||| | ||| || || |||||| |   ||||||    ||
Db    916 ACAGAACGGTGTTGATTACTGGATCGTGAAGAACTCTTGGGGAAAGAGTTGGGGTATGGA 975

Qy    998 TGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTA 1057
          ||| |  ||    || ||  | ||||| || || |    |||||| ||||| || |||
Db    976 TGGGTTTATGCACATGCAGCGTAACACCGGCAACTCGGAAGGAGTATGTGGAATCAATAT 1035

Qy   1058 TTTCGCTTCATACCCAACCAA 1078
          |||||||  || || | |||
Db   1036 GCTCGCTTCGTATCCCATCAA 1056
```

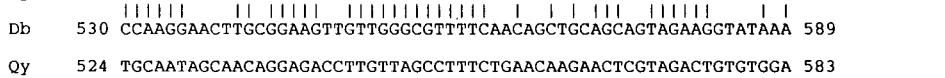
```
RESULT 5
US-09-325-932A-109
; Sequence 109, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
;  APPLICANT:  Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION:  Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of forestry plant dev
;  FILE REFERENCE:  1022
;  CURRENT APPLICATION NUMBER: US/09/325,932A
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  206
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 109
;   LENGTH: 1062
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-325-932A-109

  Query Match             11.6%;  Score 134.2;  DB 4;  Length 1062;
  Best Local Similarity   51.5%;  Pred. No. 2.4e-32;
  Matches  452;  Conservative    0;  Mismatches  393;  Indels   32;  Gaps    5;

Qy    135 CTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCA 194
            || ||||    |||| ||    ||||        || ||| |  ||  ||  || | |
Db    194 CTCTTCCGGAGATGGCAGGCGCAGCACAAGAAGGTGTACAAGCACGCCGGGGAGGCCGAG 253

Qy    195 AAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGA 254
          | | | || |||| ||||| |  |||  ||  ||  | | |||   ||         ||
Db    254 AGGCGGCTCGAGAACTTCAAAAGGAACCTGAGGTACGTGATGGA---GAGAAGCCGGAGG 310

Qy    255 AAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAG 314
          |      | |    ||| |  | ||  |||||||||| || ||| | |      | |||
Db    311 GATGGCGGCAAGCAGCATGGCGTGGGGCTGAACAAGTTCGCCGACCTGAGCAACGAGGAG 370

Qy    315 TTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAAC 374
          |||||  |    ||| ||         ||| |   |  | | || |     | ||||||
Db    371 TTCAGGCAGCGCTACCTGTCCAAGGTGAAGAAGTCGGTGAACCAGAAGTGGAGGGCCAAG 430

Qy    375 AAGAAAATG-----------AAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCA 422
          | | | |           |||| | |   |      |    | |  | || | | ||
Db    431 AGGGAGAGCTTGATGAGGAACAAGAGGAAGGGAGCGGAGTCCTGCAAGGCGCCGTCGTCT 490

Qy    423 TGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGG 482
            |||||||||||  | ||  |  |||||   || ||| |||| || |  || |||||
Db    491 CTGGATTGGAGGAACTACGGCATTGTCACCGGCGTGAAGGACCAGGGAGAATGCGGAAGT 550

Qy    483 GGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGAC 542
          | ||||||||  ||| | || |||||  || |||| |    ||||  |    |  || |||
Db    551 TGCTGGGCGTTCTCTTCGACCGGAGCAATGGAAGGAATCAATGCGCTCAAGAGCGGGGAC 610

Qy    543 CTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTAC 602
          ||  | |||||||| || ||||| ||||| ||||| |         | ||| || |   ||
Db    611 CTGATCAGCCTTTCCGAGCAAGAGCTCGTGGACTGCGACACCACCAACGACGGCTGCGAC 670

Qy    603 AATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGAT 662
            || |    | |  |  |||||| |||||  |  |  | || || || |   | ||
Db    671 GGCGGCTACATGGACTATGCGTTCGAGTGGGTCATCAACAACGGCGGTATCGATTCGGAA 730

Qy    663 GATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAG 722
          || || ||||| ||||         ||     |    |    ||  |     |  ||
Db    731 GAAGACTATCCCTACACCAGCGTCTTTGGCGAGGGTGGTATCTGCAACGTCACCAAGGAG 790

Qy    723 GTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACA 782
          |  |  |          |  |  |  |  |  ||    |  |        |  |    | |
Db    791 GAGAACA---------ACAAGGCGGTCACCATTGATGGGTATGTGGACGTTTATCCGTCG 841

Qy    783 GAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGA------T 836
          ||  | |    |  |  || |  || |  | || || ||||| |||   || ||
Db    842 GACGACGGCCTTCTCTGCACGGTCATCCAACAGCCGATTAGCGTCGGCATGGACGGCTCG 901
```

```
Qy    837 GCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCG 896
          || | |||||| ||  | ||||| ||||| || || || ||        |   | | |||
Db    902 GCGATAGATTTCCAGCTCTACACTGGGGGCATCTACGACGGTAGCTGTTCAGCGAATCCC 961

Qy    897 TATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTAC 956
          | |  ||  | ||     ||  |  | || |||||||| || |   |  || | ||||||
Db    962 GACGACATCGACCATGCGGTCCTGATAGTTGGTTATGGCTCCGAAGGCGGCGAAGATTAT 1021

Qy    957 TGGATAGCGAA--AAATTCATGGGGAGAAGATTGGGG 991
          ||||| | |||  || | | ||||||  ||||| |||
Db   1022 TGGATCGTGAAAGAACTCCTTGGGGAACAGATTTGGG 1058


RESULT 6
US-09-325-932A-98
; Sequence 98, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
;  APPLICANT:  Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION:  Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of forestry plant develo
;  FILE REFERENCE:  1022
;  CURRENT APPLICATION NUMBER: US/09/325,932A
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  206
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 98
;   LENGTH: 1600
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-325-932A-98

  Query Match             11.5%;  Score 132.8;  DB 4;  Length 1600;
  Best Local Similarity   51.0%;  Pred. No. 8.4e-32;
  Matches  466;  Conservative    0;  Mismatches  412;  Indels   36;  Gaps    5;

Qy    124 AGGTGTCTTCACTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACG 183
          |||||    |  | | | |    |||       |||| ||    | |||| |   ||  |
Db    231 AGGTGATGGCCGTCTACGAGAGCTGGCTCGCCAAGCACGGCAAGGCCTACAACGCCCTGG 290

Qy    184 AAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGA 243
          ||  |||  ||| |  |  || | |||||| | |||      |  |||   |||   |
Db    291 GCGAGAAGGAGAAGCGCTTCCAGGTCTTCAAGGACAACCTCCGGTTCATCGACGACCACA 350

Qy    244 ATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCA 303
          | ||   | |  |      ||   |  |     | ||  | ||| |||| || ||| |||
Db    351 ACGCCGGCGGGGA------CCGGACCTACACGGTCGGCCTCAACCAGTTCGCCGACCTCA 404

Qy    304 CTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCA 363
          ||    | |||| | |    | |||| ||   ||    | ||||  ||||  ||
Db    405 CTAACGAGGAGTACCGGTCCATGTACCTGGGCGCCAGGATGGATCGGTCGGGGCGGCGGC 464

Qy    364 AAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCAT 423
          |     |   |       | | |         |  | ||  | |  || || ||
Db    465 TCGGGAGGGCCCGCAGCGATCGGTACGCCGTGGCCGCCGGGGAGGAGCTGCCGGCGTCCG 524

Qy    424 GGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGG 483
          ||||||||||||  |||| | | |     | || ||| |||| ||  |||| || ||
Db    525 TCGATTGGAGGAAGGAAGGCGCCGTTGTTGACGTCAAGGACCAGGGAAGCTGCGGGAGTT 584

Qy    484 GTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACC 543
          | |||||||| ||| | |  |  ||  | || |    | |     | |  || || ||
Db    585 GCTGGGCGTTCTCTACAATTGCTGCTGTGGAGGGGATAAACAAGCTTGTGACTGGTGATT 644

Qy    544 TTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAG---AAAGCGAAGGTTCTT 600
          |  |    || || || || ||||| || ||||| |     |     | |  |||||  | |
Db    645 TGATCTCTCTGTCCGAGCAGGAACTTGTGGACTGCGATACATCCTACAATGAAGGATGTA 704

Qy    601 ACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTG 660
          |    ||      | || |  | || ||||   || |  |  | || |||||||  || |
Db    705 ATGGCGGGCTCATGGATTATGCCTTTGAATTCATTATCAACAACGGAGGCATTGATACCG 764

Qy    661 ATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATG---CAAAGCCAATAAGATACAAG 717
          | || |||||||| || ||||||  |||  | |  ||   | ||  ||  ||||    |
Db    765 AGGAAGATTATCCCTATAGAGCTGTAGATAGCACTTGTGACCAATACAGGAAGAACGCAA 824

Qy    718 ACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAG 777
          |    || || |||||||  ||||||  | |   |       |||||||
Db    825 AGGTTGTGACGATTGACGATTATGAAGATGTTCCAGAAAATGATGAGA------------ 872

Qy    778 AGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATG 837
                  |||| ||   ||  ||  |   | | |||||| | |||||  | ||||| |
Db    873 ---------AAGCATTGCAAAAGGCAGTTGCTAATCAACCAGTCAGTGTGGCCATTGAAG 923

Qy    838 CAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGT 897
          ||  ||       | ||  |        |    ||  | |  |   |   | | | ||  |
Db    924 CAGGAGGCCGGGAATTCCAGTTTTATG---ATTCGGGTATATTTACTGGCAAATGTGGGA 980

Qy    898 ATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACT 957
            |   |  ||||    |||      || || |||||  |||   | || || |||||||
Db    981 CAGCTCTGGATCATGGGGTTACTGCAGTCGGATATGGCACAGAAAACGGAGTTGATTACT 1040

Qy    958 GGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAA 1017
          |||||| ||| || |||||||| |     ||||||||  | || |||||    ||   |
Db   1041 GGATAGTGAAGAACTCATGGGGCGGTAGCTGGGGAGAGCAAGGGTACATCAAAATGGCAC 1100

Qy   1018 GAAACACGGGTAAT 1031
          ||||   ||  |||
Db   1101 GAAATGTGGCCAAT 1114


RESULT 7
US-09-325-932A-102
; Sequence 102, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
;  APPLICANT:  Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION:  Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of forestry plant dev
;  FILE REFERENCE:  1022
;  CURRENT APPLICATION NUMBER: US/09/325,932A
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  206
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 102
;   LENGTH: 2167
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-09-325-932A-102

  Query Match             11.1%;  Score 128.6;  DB 4;  Length 2167;
  Best Local Similarity   50.5%;  Pred. No. 2.1e-30;
  Matches  454;  Conservative    0;  Mismatches  424;  Indels   21;  Gaps    5;

Qy    183 GAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATG 242
          || ||  | || ||  || | ||||| ||||  | ||   ||| ||||||     | ||
Db    253 GATGAGCATGCCAAACGATTCGAGATATTCAAAGACAATGTGAAATATATCGACTCCGTG 312

Qy    243 AATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATC 302
          ||| || |  ||   ||    ||  ||      |  | | |  ||| | ||   |
Db    313 AAT-CAGAAGGATGGTCCATACAAACTTGGATTGAATAAGTTTACAGATCTGAGTAACGA 371

Qy    303 ACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATC 362
          |    |||||          |  |  ||| | | |          | | |||   ||  |
Db    372 AGAATTCAAGGCCATGCACATGACAACTAGAATGGAGAAGCACAAGAGTCTGCGCCGAGA 431

Qy    363 AAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCA 422
          | | |  ||    ||   || |   | | || | | |     |  ||| | |  || |  ||
Db    432 CAGAGGAACACACAGTGGATCATTCATGTACCAAA--ACTCCGACAACCTGCCAGAGTCT 489

Qy    423 TGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGG 482
            ||||||||  |||| || | | | |  |  || ||| |||||||    ||||||||
Db    490 ATTGATTGGAGAGAAAAGGGAGCCGTTAATCCTGTGAAGAACCAAGGCCAATGTGGAAGT 549
```

RESULT 8
US-08-821-994-62
; Sequence 62, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 62
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-62

Query Match 10.4%; Score 120.2; DB 4; Length 1434;
Best Local Similarity 51.8%; Pred. No. 7.4e-28;
Matches 477; Conservative 0; Mismatches 393; Indels 51; Gaps 7;





RESULT 9
US-08-821-994-61
; Sequence 61, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A

```
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
;   LENGTH: 1390
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-61

  Query Match             10.3%;  Score 118.6;  DB 4;  Length 1390;
  Best Local Similarity   51.7%;  Pred. No. 2.3e-27;
  Matches  476;  Conservative    0;  Mismatches  394;  Indels   51;  Gaps    7;

Qy    170 CTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTA 229
Db    233 CAACGGTATTATCAACCAACAAGATGAAAGATTCAATATTTTCAAAGACAACCTAAGATT 292

Qy    230 TATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAA 289
Db    293 CATCGATCTACACAACGAGAACAACAA---GAACGCTACTTACAAGCTTGGTCTAACCAT 349

Qy    290 GTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGT 349
Db    350 ATTCGCTGATCTCACTAACGATGAGTACCGGAGTTTATACCTCGGGGCAAGAACCGAGCC 409

Qy    350 GTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGAC-- 407
Db    410 TGTCCGCCGCATCACTAAGGCCAAGAACGTTAACATGAAATACTCAGCCGCAGTAAACGA 469

Qy    408 ----CATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTA 463
Db    470 CGTGGAGGTTCCGGAGACGGTTGACTGGAGACAGAAAGGAGCCGTTAATGCCATTAAAAA 529

Qy    464 CCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACA 523
Db    530 CCAAGGATCTTGCGGAAGTTGTTGGGCGTTTTCAACAGCTGCAGCAGTAGAAGGCATAAA 589

Qy    524 TGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGA 583
Db    590 CAAGATCGTAACAGGAGAGCTCATATCTCTGTCCGAACAAGAACTTGTCGACTGCGACAA 649

Qy    584 ---AGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGA 640
Db    650 ATCATACAACCAAGGCTGTAACGGCGGTCTAATGGATTATGCTTTTCAATTCATCATGAA 709

Qy    641 ACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAA 700
Db    710 AAACGGCGGATTAAACACCGAGCAAGACTATCCTTACCACGGAACCAATGGCAAATGCAA 769

Qy    701 AGCCAATAAGATACAAGACAAGGTT---ACAATTGACGGATATGAAACTCTAATAATGTC 757
Db    770 CTCTTTACTTAAAAATTCGAGAGTTGTGACTATCGATGGATACGAAGATGTTCCTAGTAA 829

Qy    758 AGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACC 817
Db    830 AGATGAAA----------------------CCGCGTTGAAGAGAGCAGTTTCGTACCAGCC 868

Qy    818 AATTAGTGTCTCAATTGATGC------AAAAGATTTTCATTTATACACCGGGGGAATTTA 871
Db    869 TGTGAGTGTTGCTATTGATGCTGGTGGAAGAGCTTTCCAACATTACCAATCTGGAATCTT 928

Qy    872 TGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTA 931
Db    929 CACTGGAAAGTGTGGTACGACT---------ATGGATCACGCTGTTGTGGCGGTTGGTTA 979

Qy    932 TGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGG 991
Db    980 TGGATCAGAGAACGGTGTTGACTATTGGATTGTACGTAACTCTTGGGGTACAAGCTGGGG 1039

Qy    992 AGAAGATGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGAT 1051
Db   1040 AGAGGATGGTTACATTAGGATGGAGAGAAACGTGGCGTCCAAATCCGGTAAGTGTGGGAT 1099

Qy   1052 GAATTATTTCGCTTCATACCC 1072
Db   1100 TGCGATTGAAGCCTCGTATCC 1120


RESULT 10
US-08-821-994-63
; Sequence 63, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-63

  Query Match             10.3%;  Score 118.6;  DB 4;  Length 1441;
  Best Local Similarity   51.7%;  Pred. No. 2.4e-27;
  Matches  476;  Conservative    0;  Mismatches  394;  Indels   51;  Gaps    7;

Qy    170 CTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAGATTTTCAAGAATAACTCGAACTA 229
Db    219 CAACGGTATTATCAACCAACAAGATGAAAGATTCAATATTTTCAAAGACAACCTAAGATT 278

Qy    230 TATCAGGGACATGAATGCAAACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAA 289
Db    279 CATCGATCTACACAACGAGAACAACAA---GAACGCTACTTACAAGCTTGGTCTAACCAT 335

Qy    290 GTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGT 349
Db    336 ATTCGCTGATCTCACTAACGATGAGTACCGGAGTTTATACCTCGGGGCAAGAACCGAGCC 395

Qy    350 GTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGAC-- 407
Db    396 TGTCCGCCGCATCACTAAGGCCAAGAACGTTAACATGAAATACTCAGCCGCAGTAAACGA 455

Qy    408 ----CATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTA 463
Db    456 CGTGGAGGTTCCGGAGACGGTTGACTGGAGACAGAAAGGAGCCGTTAATGCCATTAAAAA 515

Qy    464 CCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACA 523
Db    516 CCAAGGATCTTGCGGAAGTTGTTGGGCGTTTTCAACAGCTGCAGCAGTAGAAGGCATAAA 575

Qy    524 TGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGA 583
Db    576 CAAGATCGTAACAGGAGAGCTCATATCTCTGTCCGAACAAGAACTTGTCGACTGCGACAA 635

Qy    584 ---AGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGA 640
Db    636 ATCATACAACCAAGGCTGTAACGGCGGTCTAATGGATTATGCTTTTCAATTCATCATGAA 695

Qy    641 ACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAA 700
Db    696 AAACGGCGGATTAAACACCGAGCAAGACTATCCTTACCACGGAACCAATGGCAAATGCAA 755
```

```
Qy    701 AGCCAATAAGATACAAGACAAGGTT---ACAATTGACGGATATGAAACTCTAATAATGTC 757
Db    756 CTCTTTACTTAAAAATTCGAGAGTTGTGACTATCGATGGATACGAAGATGTTCCTAGTAA 815

Qy    758 AGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACC 817
Db    816 AGATGAAA----------CCGCGTTGAAGAGAGCAGTTTCGTACCAGCC 854

Qy    818 AATTAGTGTCTCAATTGATGC------AAAAGATTTTCATTTATACACCGGGGGAATTTA 871
Db    855 TGTGAGTGTTGCTATTGATGCTGGTGGAAGAGCTTTCCAACATTACCAATCTGGAATCTT 914

Qy    872 TGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTA 931
Db    915 CACTGGAAAGTGTGGTACGACT---------ATGGATCACGCTGTTGTGGCGGTTGGTTA 965

Qy    932 TGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGG 991
Db    966 TGGATCAGAGAACGGTGTTGACTATTGGATTGTACGTAACTCTTGGGGTACAAGCTGGGG 1025

Qy    992 AGAAGATGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGAT 1051
Db   1026 AGAGGATGGTTACATTAGGATGGAGAGAAACGTGGCGTCCAAATCCGGTAAGTGTGGGAT 1085

Qy   1052 GAATTATTTCGCTTCATACCC 1072
Db   1086 TGCGATTGAAGCCTCGTATCC 1106
```

RESULT 11
US-09-325-932A-110
; Sequence 110, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant develo
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-110

Query Match 10.1%; Score 117.2; DB 4; Length 765;
Best Local Similarity 53.6%; Pred. No. 4.5e-27;
Matches 298; Conservative 0; Mismatches 243; Indels 15; Gaps 2;

```
Qy    135 CTGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCA 194
Db    197 CTCTTCCGGAGATGGCAGGCGCAGCACAAGAAGGTGTACAAGCACGCCGGGGAGGCCGAG 256

Qy    195 AAGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGA 254
Db    257 AGGCGGCTCGAGAACTTCAAAAGGAACCTGAGGTACGTGATGGA---GAGAAGCCGGAGG 313

Qy    255 AAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAG 314
Db    314 GATGGCGGCAAGCAGCATGGCGTGGGGCTGAACAAGTTCGCCGACCTGAGCAACGAGGAG 373

Qy    315 TTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAAC 374
Db    374 TTCAGGCAGCTCTACCTGTCCAAGGTGAAGAAGTCGGTGAACCAGAAGTGGAGGGCCAAG 433

Qy    375 AAGAAAATG------------AAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCA 422
Db    434 AGGGAGAGCTTGATGAGGAACAAGAGGAAGGGAGCGGAGTCCTGCAAGGCGCCGTCGTCT 493
```

```
Qy    423 TGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGG 482
Db    494 CTGGACTGGAGGAACTACGGCATTGTCACCGGCGTGAAGGACCAGGGAGAATGCGGAAGT 553

Qy    483 GGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGAC 542
Db    554 TGCTGGGCGTTCTCTTCGACCGGAGCAATGGAAGGAATCAATGCGCTCAAGAGCGGTGAC 613

Qy    543 CTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTAC 602
Db    614 CTGATCAGCCTTTCCGAGCAAGAGCTCGTGGACTGCGACACCACCAACGACGGGTGCGAC 673

Qy    603 AATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGAT 662
Db    674 GGCGGCTACATGGACTATGCCTTCGAGTGGGTCATCAACAACGGCGGTATCGATTCGGAA 733

Qy    663 GATGATTATCCTTACA 678
Db    734 GAAGACTATCCCTACA 749
```

RESULT 12
US-08-821-994-86
; Sequence 86, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
; LENGTH: 1102
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-86

Query Match 10.1%; Score 117; DB 4; Length 1102;
Best Local Similarity 54.9%; Pred. No. 6.4e-27;
Matches 362; Conservative 0; Mismatches 255; Indels 42; Gaps 5;

```
Qy    426 GATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGT 485
Db    315 GACTGGAGACAGAAAGGAGCCGTTAATGCCATTAAAGACCAAGGAACTTGCGGAAGTTGT 374

Qy    486 TGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTT 545
Db    375 TGGGCGTTTTCAACAGCTGCAGCAGTAGAAGGTATAAACAAGATCGTAACAGGAGAACTC 434

Qy    546 GTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAA---GAAAGCGAAGGTTCTTAC 602
Db    435 GTATCTTTGTCCGAACAAGAACTTGTCGACTGCGACAAATCGTACAACCAAGGCTGTAAC 494

Qy    603 AATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGAT 662
Db    495 GGCGGTCTAATGGATTATGCTTTTCAATTCATAATGAAAAACGGCGGATTAAACACCGAG 554

Qy    663 GATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGC---CAATAAGATACAAGAC 719
Db    555 AAAGACTATCCTTACCACGGAACCAATGGCAAATGCAACTCTTTACTTAAGAATTCAAGA 614

Qy    720 AAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAG 779
Db    615 GTTGTAACTATCGATGGATACGAAGATGTTCCTAGTAAAGATGAAA---------- 660
```

```
Qy    780 ACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGC- 838
Db    661 -------CCGCGTTGAAGAGAGCAGTTTCATACCAGCCTGTGAGTGTTGCTATTGATGCT 713

Qy    839 -----AAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGT 893
Db    714 GGTGGAAGAGCTTTCCAACATTACCAATCTGGAATCTTCACTGGAAAGTGTGGTACGAAT 773

Qy    894 CCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGAT 953
Db    774 ---------ATGGATCACGCTGTGGTGGCGGTTGGTTATGGGTCAGAGAACGGCGTTGAC 824

Qy    954 TACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATC 1013
Db    825 TATTGGATTGTACGTAACTCTTGGGGTACACGTTGGGGAGAGGATGGTTACATTAGGATG 884

Qy   1014 CAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCC 1072
Db    885 GAGAGAAACGTGGCGTCTAAATCCGGTAAGTGTGGGATTGCGATAGAAGCCTCGTATCC 943


RESULT 13
US-08-453-924-9
; Sequence 9, Application US/08453924
; Patent No. 5608152
;  GENERAL INFORMATION:
;    APPLICANT:  Kridl, Jean C.
;    APPLICANT:  Knauf, Vic C.
;    TITLE OF INVENTION:  Seed-Specific Transcriptional Regulation
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Weil, Gotshal & Manges
;      STREET:  2882 Sand Hill Road, Ste. 280
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,924
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/742,834
;      FILING DATE:  08-AUGUST-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/550,804
;      FILING DATE:  09-JULY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/147,781
;      FILING DATE:  25-JANUARY-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/078,538
;      FILING DATE:  28-JULY-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  06/891,529
;      FILING DATE:  31-JULY-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rae-Venter, Barbara
;      REGISTRATION NUMBER:  32,750
;      REFERENCE/DOCKET NUMBER:  CGNE-037/04US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 926-6200
;      TELEFAX:  (415) 854-3713
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1186 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..1051
US-08-453-924-9

  Query Match             10.1%;  Score 117;  DB 1;  Length 1186;
  Best Local Similarity   47.3%;  Pred. No. 6.7e-27;
  Matches  440;  Conservative    0;  Mismatches  470;  Indels   21;  Gaps    2;

Qy    147 TGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAAAGAGACTTGAG 206
Db    140 TGGATGACCGAACACGGCCGTGTTTACGCAGATGCGAACGAGAAAAACAACCGCTACGCT 199

Qy    207 ATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCCAT 266
Db    200 GTTTTCAAACGCAACGTGGAACGCATTGAACGCTTAAATG---ACGTTCAATCCGGACTA 256

Qy    267 TCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAAAG 326
Db    257 ACGTTTAAACTCGCGGTGAACCAGTTTGCTGATCTAACCAACGAAGAATTCCGTTCTATG 316

Qy    327 TACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACAAGAAAATGAAG 386
Db    317 TACACTGGTTTCAAAGGAAACTCTGTGTTGTCTAGTCGAACTAAACCAACGTCGTTTAGG 376

Qy    387 AAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTC 446
Db    377 TACCAAAACGTTTCTTCTGATGCGTTGCCGGTTTCTGTTGATTGGAGGAAGAAAGGAGCT 436

Qy    447 ATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGA 506
Db    437 GTGACTCCTATCAAGGATCAAGGCTTATGCGGATCTTGTTGGGCGTTTTCAGCTGTTGCG 496

Qy    507 GCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAA 566
Db    497 GCTATAGAAGGAGTAGCACAGATAAAGAAAGGGAAACTCATTTCTTTGTCTGAACAAGAG 556

Qy    567 CTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTC 626
Db    557 CTTGTCGACTGCGACACAAACGATGGTGGCTGCATGGGCGGTTTGATGGATACAGCGTTT 616

Qy    627 GAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAA 686
Db    617 AACTACACAATAACTATTGGCGGCTTAACCTCTGAATCAAATTATCCTTATAAAAGCACA 676

Qy    687 GAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACT 746
Db    677 AACGGCACTTGCAACTTCAATAAAACTAAACAGA----------------TAGCAACT 718

Qy    747 CTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATC 806
Db    719 TCTATCAAAGGTTTTGAGGATGTCCCGGCTAACGATGAGAAAGCCCTAATGAAGGCAGTG 778

Qy    807 CTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGA 866
Db    779 GCACACCACCCGGTTAGCATTGGAATAGCGGGAGGAGATATTGGTTTCCAATTCTATTCG 838

Qy    867 ATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTG 926
Db    839 TCCGGTGTGTTCAGCGGAGAATGCACAACTCATCTTGATCACGGGGTAACTGCGGTTGGA 898

Qy    927 GGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGAT 986
Db    899 TACGGCCGATCTAAAAACGGATTAAAGTACTGGATCCTCAAGAATTCATGGGGACCAAAA 958

Qy    987 TGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGT 1046
Db    959 TGGGGAGAACGTGGATACATGAGGATCAAAAAAGATATCAAGCCTAAACACGGACAATGT 1018

Qy   1047 GGGATGAATTATTTCGCTTCATACCCAACCA 1077
```

```
Db   1019 GGTCTTGCCATGAATGCTTCGTACCCAACTA 1049


RESULT 14
US-08-821-994-82
; Sequence 82, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
;   LENGTH: 1661
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 4..10, 12, 90, 428, 775, 776, 1079..1081, 1134, 1148..1165
;   OTHER INFORMATION: n is uncertain
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1172, 1180, 1191..1198, 1213..1226, 1250..1253, 1274, 1277
;   OTHER INFORMATION: n is uncertain
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1279..1282, 1291..1303, 1322..1325, 1339, 1340, 1345..1348
;   OTHER INFORMATION: n is uncertain
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1357..1371, 1391..1410, 1420, 1427..1435, 1448..1453, 1459..1464
;   OTHER INFORMATION: n is uncertain
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1487, 1501..1504, 1512, 1513, 1529..1538, 1546..1550, 1556
;   OTHER INFORMATION: n is uncertain
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1558..1562, 1564..1566, 1568, 1570, 1574..1579, 1582, 1583, 1586
;   OTHER INFORMATION: n is uncertain
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1590..1594, 1615, 1638..1647
;   OTHER INFORMATION: n is uncertain
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Majority
US-08-821-994-82

  Query Match            10.0%;  Score 115.2;  DB 4;  Length 1661;
  Best Local Similarity   53.3%;  Pred. No. 3e-26;
  Matches  322;  Conservative    0;  Mismatches  260;  Indels   22;  Gaps    3;

Qy    426 GATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGT 485
          || |||||  | |||||  | | | |      | ||  |||||||    || |||||  ||
Db    493 GACTGGAGACAGAAAGGAGCCGTTAATGCCATTAAAGACCAAGGAACTTGCGGAAGTTGT 552

Qy    486 TGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTT 545
          ||||||||||  |  | | |||  ||||||    | |    || | |||||||| ||
Db    553 TGGGCGTTTTCAACAGCTGCAGCAGTAGAAGGTATAAACAAGATCGTAACAGGAGAACTC 612

Qy    546 GTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGA---AGAAAGCGAAGGTTCTTAC 602
           |    || || |||||||||| || |||||  |    |  | | | | |||| | | ||
Db    613 ATATCTCTGTCCGAACAAGAACTTGTCGACTGCGACAAATCATACAACCAAGGCTGTAAC 672

Qy    603 AATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGAT 662
             ||      | || |  | ||  |||    |  |  || | || ||  |   ||| ||
Db    673 GGCGGTCTAATGGATTATGCTTTTCAATTCATCATGAAAAACGGCGGATTAAACACCGAG 732

Qy    663 GATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAG 722
          | || |||||||||   |  |   | || | ||||||  |   | |  || ||    |
Db    733 AAAGACTATCCTTACCACGGAACCAATGGCAAATGCAACTCTNNTTACTTAAAAATTCAA 792

Qy    723 GTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACA 782
          |              || ||    ||  |   ||   |||  | |        | |
Db    793 GAGTTGTGACTATCGATGGATACGAAGATGTTCCTAGTAAAGATGAAACCGCGTTGAAGA 852

Qy    783 GAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAA 842
          |||||          |||       |   || |   ||            || || || |
Db    853 GAGCAGTTTCGTACCAGCCTGTGAGTGTTGCTATTGATG----------CTGGTGGAAGA 902

Qy    843 GATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGG 902
          | ||| ||    |||      ||||| |    |||| |     |||| | |
Db    903 GCTTTCCAACATTACCAATCTGGAATCTTCACTGGAAAGTGTGGTACGACT--------- 953

Qy    903 ATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATA 962
          ||  |||||  |||  |    || |||||||||||| | | ||||| || || |||||
Db    954 ATGGATCACGCTGTGGTGGCGGTTGGTTATGGTTCAGAGAACGGTGTTGACTATTGGATT 1013

Qy    963 GCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAAC 1022
          |     || || |||||   |  |||||||| |||||||||||| ||||  | ||||||
Db   1014 GTACGTAACTCTTGGGGTACAAGTTGGGGAGAGGATGGTTACATTAGGATGGAGAGAAAC 1073

Qy   1023 ACGG 1026
            ||
Db   1074 GTGG 1077


RESULT 15
US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
;   LENGTH: 1474
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-64

  Query Match             9.9%;  Score 114.8;  DB 4;  Length 1474;
  Best Local Similarity   53.5%;  Pred. No. 3.8e-26;
  Matches  323;  Conservative    0;  Mismatches  257;  Indels   24;  Gaps    3;

Qy    426 GATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGT 485
          ||||||||    ||||| | | | |      | ||  |||||||    || |||||  ||
Db    480 GATTGGAGACGGAAAGGAGCCGTTAATGCCATTAAAAACCAAGGAACTTGCGGAAGTTGT 539

Qy    486 TGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTT 545
          ||||||||||  |  | | |||  ||||||    | |    || | |||||||| ||
Db    540 TGGGCGTTTTCGACAGCTGCAGCAGTAGAAGGTATAAACAAGATCGTAACAGGAGAACTC 599

Qy    546 GTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAA---GAAAGCGAAGGTTCTTAC 602
           |    || || |||||||||| || ||||| |    |    | | | |||| |   ||
Db    600 ATATCTCTGTCCGAACAAGAACTTGTCGACTGCGACAGATCCTACAACCAAGGCTGCAAC 659
```

```
Qy     603 AATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGAT 662
           ||| |    | |  |  | ||  |||   |  |  || | || ||  |   ||| ||
Db     660 GGTGGTTTAATGGACTATGCTTTTCAATTCATCATGAAAAACGGCGGTTTGAACACCGAG 719

Qy     663 GATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAG 722
            | ||||||||||||| | | |    | || | ||||||  |      ||   |    |
Db     720 CAAGATTATCCTTACCGTGGTTCCAATGGAAAATGCAATTCTTTACTGAAGAATTCAAGA 779

Qy     723 GTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACA 782
           |||  || |   |    | |        |     | || ||    || |  |   ||| |
Db     780 GTTGTAACTATTGATGGTTACGAAGATGTTCCTACTGAAGATGAAACGGCGTTGAAGAGA 839

Qy     783 GAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAA 842
           |      | | |     ||  |   |||               |   |  || || || |
Db     840 GCAGTTTCATACCAGCCCGTGAGTGTTG----------CCATTGAAGCTGGTGGAAGA 887

Qy     843 GATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGG 902
           | ||| ||    |||     ||| || |    |||| |     | |||| ||
Db     888 GTTTTCCAACATTACCAATCGGGGATCTTCACTGGAAAGTGTGGGACAAATC-------- 939

Qy     903 ATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATA 962
           |  ||||    ||  |   ||| |||||||||||||| | | ||| | || || |||||
Db     940 -TAGATCATGCAGTGGTGGCTGTTGGTTATGGTTCAGAGAACGGTATTGACTATTGGATT 998

Qy     963 GCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAAC 1022
           |  |  || || |||||   |  |||||||| ||||||||||| ||||  | ||||||
Db     999 GTAAGGAACTCGTGGGGTACACGTTGGGGAGAGGATGGTTACATTAGGATGGAGAGAAAC 1058

Qy    1023 ACGG 1026
             ||
Db    1059 TTGG 1062
```

Search completed: January 11, 2003, 21:11:10
Job time : 61 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 20:37:43 ; Search time 64 Seconds
(without alignments)
7949.367 Million cell updates/sec

Title: US-09-805-694B-1
Perfect score: 1156
Sequence: 1 gcggccgcatgggtttcctt.....tctcctctttgagcggccgc 1156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	144	12.5	1392	9	US-09-938-842A-1372	Sequence 1372, Ap
	2	128.6	11.1	1116	9	US-09-938-842A-2449	Sequence 2449, Ap
	3	117.6	10.2	1086	9	US-09-938-842A-2484	Sequence 2484, Ap
	4	103.2	8.9	385	10	US-09-878-574-140	Sequence 140, App
	5	102.4	8.9	1047	9	US-09-938-842A-742	Sequence 742, App
	6	102.4	8.9	1047	10	US-09-887-576-668	Sequence 668, App
c	7	95.8	8.3	1282	10	US-09-770-445-20	Sequence 20, Appl
	8	86.8	7.5	1095	9	US-09-938-842A-2351	Sequence 2351, Ap
	9	85.6	7.4	390	10	US-09-878-574-1099	Sequence 1099, Ap
c	10	80.8	7.0	604	10	US-09-770-149-933	Sequence 933, App
	11	74	6.4	1122	9	US-09-938-842A-1324	Sequence 1324, Ap
	12	72.4	6.3	264	10	US-09-878-574-10000	Sequence 10000, A
	13	72.2	6.2	1643	9	US-10-099-275-1	Sequence 1, Appli
	14	71.2	6.2	2004	10	US-09-887-576-294	Sequence 294, App
	15	68.2	5.9	996	9	US-09-990-064-1	Sequence 1, Appli
	16	68	5.9	1642	9	US-10-094-080-2	Sequence 2, Appli
	17	67.8	5.9	1405	10	US-09-943-272-1	Sequence 1, Appli
	18	64.6	5.6	1619	10	US-09-953-956-1	Sequence 1, Appli
	19	64.6	5.6	1619	12	US-10-114-464-1	Sequence 1, Appli
	20	64.6	5.6	1671	12	US-10-044-090-462	Sequence 462, App
	21	64.6	5.6	1756	10	US-09-925-300-753	Sequence 753, App
	22	64.4	5.6	269	10	US-09-878-574-12494	Sequence 12494, A
	23	62.2	5.4	1490	10	US-09-925-302-35	Sequence 35, Appl
	24	62	5.4	1486	10	US-09-728-952-34	Sequence 34, Appl
	25	62	5.4	1486	10	US-09-728-952-57	Sequence 57, Appl
	26	61.4	5.3	768	10	US-09-910-943-32	Sequence 32, Appl
	27	58.6	5.1	1107	9	US-09-938-842A-1386	Sequence 1386, Ap
	28	58	5.0	1369	9	US-10-076-785-23	Sequence 23, Appl
c	29	57.4	5.0	522	10	US-09-924-035A-97	Sequence 97, Appl
	30	57	4.9	1002	10	US-09-529-063-2	Sequence 2, Appli
	31	57	4.9	1370	10	US-09-529-063-3	Sequence 3, Appli
	32	57	4.9	1547	9	US-10-028-072-11	Sequence 11, Appl
	33	57	4.9	1790	10	US-09-917-800A-1441	Sequence 1441, Ap
	34	56.4	4.9	743	9	US-10-076-785-21	Sequence 21, Appl
	35	55.6	4.8	390	10	US-09-960-352-4092	Sequence 4092, Ap
	36	54	4.7	250	10	US-09-923-876-6094	Sequence 6094, Ap
	37	53.8	4.7	364	9	US-09-796-692-5116	Sequence 5116, Ap
	38	53.2	4.6	266	10	US-09-923-876-5898	Sequence 5898, Ap
c	39	53	4.6	358	10	US-09-998-598-2051	Sequence 2051, Ap
	40	52	4.5	292	10	US-09-878-574-2624	Sequence 2624, Ap
	41	51.8	4.5	273	10	US-09-878-574-11726	Sequence 11726, A
	42	51.6	4.5	1017	10	US-09-800-528-7	Sequence 7, Appli
	43	50.8	4.4	332	10	US-09-923-876-3606	Sequence 3606, Ap
	44	50.8	4.4	505	9	US-10-076-785-13	Sequence 13, Appl
	45	49.8	4.3	2251	10	US-09-925-301-179	Sequence 179, App

ALIGNMENTS

RESULT 1
US-09-938-842A-1372
; Sequence 1372, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1372
; LENGTH: 1392
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1372

Query Match 12.5%; Score 144; DB 9; Length 1392;
Best Local Similarity 54.2%; Pred. No. 3.7e-32;
Matches 372; Conservative 0; Mismatches 290; Indels 24; Gaps 3;

```
Qy    401 TTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAA 460
          | |||||   |  ||||  ||    || ||||| ||| |||||| | || |  | || ||
Db    402 TGGTGACGCGCTTCCTGATTCGGTTGACTGGAGAAAAGAAGGTGCCGTCGCTGATGTTAA 461

Qy    461 GTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGC 520
          | |||||  ||||||| ||  |||||| || ||  | |  |||||  | |||| |
Db    462 AGATCAAGGAAGCTGTGGGAGTTGTTGGGCATTCTCAACTATTGGAGCAGTGGAAGGAAT 521

Qy    521 ACATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGT 580
          | |    || |  |||||||  |  |     | ||||||||||| | || || ||||
Db    522 AAACAAGATTGTGACAGGAGATTTAATATCTTTGTCTGAACAAGAATTGGTTGATTGTGA 581
```

```
Qy    581 GGAA---GAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTT 637
Db    582 CACATCTTATAACCAAGGATGTAATGGTGGTTTAATGGACTATGCGTTTGAGTTTATTAT 641

Qy    638 AGAACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATG 697
Db    642 CAAAAATGGTGGTATTGATACCGAAGCGGATTATCCTTACAAGGCTGCTGATGGTCGTTG 701

Qy    698 CAAAGCCAATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTC 757
Db    702 TGACCAGAACAGGAAAAATGCTAAGGTT----------------GTTACCATTGATTC 743

Qy    758 AGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACC 817
Db    744 ATATGAAGATGTCCCTGAGAACAGCGAGGCGTCTCTAAAGAAAGCTTTGGCTCACCAGCC 803

Qy    818 AATTAGTGTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGG 877
Db    804 TATTAGTGTCGCCATTGAAGCGGGTGGTCGCGCTTTCCAGCTC---TATTCTTCGGGTGT 860

Qy    878 AGAAAACTGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTC 937
Db    861 ATTTGATGGACTTTGTGGGACGGAGTTAGACCACGGTGTTGTGGCTGTGGGTTATGGAAC 920

Qy    938 AGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA 997
Db    921 CGAGAACGGAAAGGACTACTGGATTGTGAGAAACTCATGGGGAAACAGATGGGGAGAGAG 980

Qy    998 TGGTTACATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTA 1057
Db    981 CGGATACATAAAGATGGCACGTAACATTGAAGCGCCAACAGGAAAGTGTGGAATCGCTAT 1040

Qy   1058 TTTCGCTTCATACCCAACCAAAGAGG 1083
Db   1041 GGAGGCTTCGTATCCGATCAAGAAAG 1066
```

RESULT 2
US-09-938-842A-2449
; Sequence 2449, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2449
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2449

Query Match 11.1%; Score 128.6; DB 9; Length 1116;
Best Local Similarity 49.6%; Pred. No. 1.1e-27;
Matches 471; Conservative 0; Mismatches 439; Indels 39; Gaps 4;

```
Qy    136 TGTTCCAACTATGGAAGAGTGAGCATGGACGTGTCTACCATAACCACGAAGAAGAGGCAA 195
Db    164 TGTTCGAGTCATGGATGGTCAAACATGGGAAAGTGTACGACTCCGTTGCCGAGAAGGAAC 223

Qy    196 AGAGACTTGAGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAA 255
Db    224 GGCGTTTAACTATTTTTGAGGACAACCTCCGTTTTATCACTAACCGGAACGCTGAGA--- 280

Qy    256 AATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGT 315
Db    281 ------ATCTCAGTTACCGGCTTGGTTTGAATCGGTTTGCGGATTTATCTCTCCATGAAT 334

Qy    316 TCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGCAGCAAATCAAAATGGCCAACA 375
Db    335 ACGGAGAAATTTGCCATGGGGCTGATCCAAGACCTCCTAGGAACCACGTCTTCATGACTA 394

Qy    376 AGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGA 435
Db    395 GCAGCAACCGATACAAGACTAGTGATGGTGATGTTCTTCCTAAGTCCGTTGACTGGAGAA 454

Qy    436 AAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTT 495
Db    455 ACGAAGGCGCAGTGACTGAAGTCAAAGATCAAGGCCTTTGCAGGAGTTGTTGGGCTTTCT 514

Qy    496 CTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTT 555
Db    515 CCACTGTGGGAGCAGTGGAAGGCTTAAACAAGATTGTGACTGGAGAGCTAGTAACTTTGT 574

Qy    556 CTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGT 615
Db    575 CTGAGCAAGATTTGATCAATTGTAACAAAGAAAACAATGGTTGCGGAGGAGGCAAAGTCG 634

Qy    616 ATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATCCTT 675
Db    635 AGACAGCCTATGAGTTCATCATGAACAATGGTGGTCTTGGTACCGACAACGATTATCCTT 694

Qy    676 ACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTGACG 735
Db    695 ACAAAGCTCTCAATGGAGTCTGCGAAGGCCGCCTCAAGGAAGACAA-------------- 740

Qy    736 GATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCT 795
Db    741 -CAAGAATGTTATGATTGATGGGTATGAGAATTTGCCTGCAAACGATGAAGCCGCTCTCA 799

Qy    796 TAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAA------AAGATTTTC 849
Db    800 TGAAAGCGGTTGCTCACCAGCCTGTGACTGCCGTTGTCGATTCCAGCAGCCGAGAGTTTC 859

Qy    850 ATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATC 909
Db    860 AGCTTTATGAATCGGGAGTGTTTGACGGAACTTGCGGAACAAACC---------TAAACC 910

Qy    910 ACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAA 969
Db    911 ATGGTGTTGTTGTGGTCGGGTATGGAACCGAGAATGGTCGTGACTACTGGATTGTGAAAA 970

Qy    970 ATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGTA 1029
Db    971 ACTCGAGGGGCGACACATGGGGGGAGGCTGGCTACATGAAGATGGCTCGCAACATTGCCA 1030

Qy   1030 ATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAA 1078
Db   1031 ATCCAAGAGGCCTATGTGGCATCGCAATGCGAGCTTCATACCCTCTCAA 1079
```

RESULT 3
US-09-938-842A-2484
; Sequence 2484, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3

```
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2484
;   LENGTH: 1086
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-2484

  Query Match            10.2%;  Score 117.6;  DB 9;  Length 1086;
  Best Local Similarity   49.5%;  Pred. No. 1.8e-24;
  Matches  439;  Conservative    0;  Mismatches  414;  Indels   33;  Gaps    4;

Qy    205 AGATTTTCAAGAATAACTCGAACTATATCAGGGACATGAATGCAAACAGAAAATCACCCC 264
          ||| |||||   |  |    || || ||| | |  |  |    | ||  ||      ||
Db    167 AGAGATTCAACGTTTTCAGACACAATGTCATGCATGTCCACAATACCAACAAGAAGAACC 226

Qy    265 ATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCTCAAGAGTTCAGCAAAA 324
            ||  |      |     | ||||||||||| ||  | ||   | | |||||||  ||
Db    227 GATCCTACAAACTTAAGCTCAACAAGTTTGCAGATTTAACAATTAATGAGTTCAAGAATG 286

Qy    325 AGTACTTGCAAGCTCCCA---------AGGATGTGTCGCAGCAAATCAAAATGGCCAACA 375
            |||       |   ||         |     ||| |||   |   |||   | |   |
Db    287 CATACACCGGTTCCAACATCAAGCATCACAGAATGTTGCAAGGACCAAAACGCGGCTCAA 346

Qy    376 AGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGATTGGAGGA 435
          |  |  | | | | || ||  |   ||    | |   |||  | ||    |||||||| |
Db    347 AACAGTTCATGTATGATCATGAGAATTTATCCAAATTACCGTCCTCCGTTGATTGGAGAA 406

Qy    436 AAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTT 495
          | |||||||   ||||  || | ||| | |||||    ||||||||  |||||||  || |
Db    407 AGAAAGGTGCTGTCACTGAAATCAAGAATCAAGGAAAATGTGGAAGTTGTTGGGCATTCT 466

Qy    496 CTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTTAGCCTTT 555
          || ||   | |||  | || | |     |    ||    ||      | || ||    || |
Db    467 CTACCGTAGCAGCTGTTGAGGGAATCAACAAGATCAAGACTAACAAACTGGTCTCACTGT 526

Qy    556 CTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGCGAAGGTTCTTACAATGGATGGC 612
          |||||||||||||  || || ||||      |  | |  ||||||| | ||   ||
Db    527 CTGAACAAGAACTAGTGGATTGTGATACTAAGCAGAATGAAGGTTGTAACGGAGGTCTTA 586

Qy    613 AGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGATTATC 672
          | |     | || ||||   |     |  | ||||| ||  |||||||  |||   ||||
Db    587 TGGAGATTGCATTTGAATTCATCAAGAAGAACGGTGGTATCACCACTGAAGATAGCTATC 646

Qy    673 CTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACAATTG 732
          ||||   || ||  || || | |||| | ||    |||    | | |  | ||    |
Db    647 CTTATGAAGGTATCGATGGAAAATGCGACGCTTCAAAGGATAACGGCGTGCTTGTGA--- 703

Qy    733 ACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGT 792
                      |||       ||||   |  |     | |    ||  | ||
Db    704 ---------------CGATAGATGGGCATGAAGATGTACCCGAAAACGACGAAAATGCTC 748

Qy    793 TCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAAAGATTTTCATT 852
          |  | |  ||  |     | |||||  |   |||  ||||||||||   |  ||   |||
Db    749 TTCTTAAGGCTGTTGCAAACCAACCTGTATCTGTTGCAATTGATGCTGGAAGTTCAGATT 808

Qy    853 TATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGGGATTAATCACT 912
          |  |          |    | |  |      |   |  |  ||      | | | || ||
Db    809 TCCAGTTTTATTCA---GAGGGAGTGTTTACGGGATCTTGTGGAACAGAGCTGAACCATG 865

Qy    913 TTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATT 972
          |||   |   || || ||||| || | |   ||   | | || |||||||  || ||| |
Db    866 GTGTGGCAGCGGTCGGGTATGGATCTGAGAGGGGGAAAAAGTATTGGATAGTGAGAAACT 925

Qy    973 CATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGTAATT 1032
          |||||||||  || ||||||||  |  || || ||   |||  ||||| | |  | | |
Db    926 CATGGGGAGCTGAATGGGGAGAGGGAGGCTATATAAAGATTGAAAGAGAGATTGATGAAC 985

Qy   1033 TATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACCAA 1078
               |||    || || ||         || || || || | |||
Db    986 CGGAAGGGCGTTGCGGTATTGCGATGGAGGCATCGTATCCGATCAA 1031


RESULT 4
US-09-878-574-140
; Sequence 140, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 140
;   LENGTH: 385
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(385)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3028-054-Q1-B1-H10
US-09-878-574-140

  Query Match              8.9%;  Score 103.2;  DB 10;  Length 385;
  Best Local Similarity   60.5%;  Pred. No. 1.6e-20;
  Matches  187;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy    403 GTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGT 462
          | ||| |    ||||  ||    ||||||||||  ||||||   |  | |  || ||
Db     77 GCGACAAATTGCCTGATTCCGTTGATTGGAGGAAGGAAGGTGCTGTTCCTCCTGTCAAAG 136

Qy    463 ACCAAGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCAC 522
          ||||||| |||||||| ||  |||||||  || || || |  || ||  ||||||  |  |
Db    137 ACCAAGGAGGCTGTGGGAGCTGTTGGGCATTCTCAGCAATCGGTGCAGTAGAAGGAATAA 196

Qy    523 ATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG--- 579
          ||   |||| |||||| || ||  ||    | || |||||||||  | || || ||||
Db    197 ATAAGATAGTAACAGGCGAACTGATTTCGTTATCAGAACAAGAATTGGTGGATTGTGATA 256

Qy    580 TGGAAGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAG 639
            | | | | |||| |   |    |||     | |  |  | || || |   |  |
Db    257 CTGGATATAACCAAGGATGCAATGGAGGACTTATGGACTATGCATTTGAGTTCATAATCA 316

Qy    640 AACATGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCA 699
          |  |||| |||||||   |||||||  ||||| ||      | | |   || |||||||||
Db    317 ACAATGGCGGCATTGATTCTGATGAGGATTACCCNATACGTGGTGTTGATGGTAGATGCG 376

Qy    700 AAGCCAATA 708
          |  |  |||
Db    377 ACACATATA 385


RESULT 5
US-09-938-842A-742
; Sequence 742, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
```

```
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 742
;   LENGTH: 1047
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-742

  Query Match             8.9%;  Score 102.4;  DB 9;  Length 1047;
  Best Local Similarity   49.7%;  Pred. No. 4.9e-20;
  Matches  324;  Conservative    0;  Mismatches  316;  Indels   12;  Gaps    2;

Qy    425 GGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGG 484
          |||||||||  || | || |    | || |  || ||||| |||||  | ||||||  |  |
Db    396 GGATTGGAGACAAGAGGGGGCTGTTACACCCGTTAAGTATCAAGGCCGATGTGGAGGATG 455

Qy    485 TTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCT 544
          ||||||||||| ||   ||  ||  | ||||           ||  | | ||| || ||
Db    456 CTGGGCGTTTTCCGCGGTGGCGGCAGTGGAAGGCATTACAAAGATTACTAAAGGCGAGCT 515

Qy    545 TGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGCGAAGGTTCTTA 601
          |||    || || || ||| ||||| |||||||  |     ||  | |   ||||  | |
Db    516 TGTATCGCTATCGGAGCAACAACTCTTAGACTGCGACAGAGACTACAATCAAGGATGTCG 575

Qy    602 CAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGA 661
            ||      ||   || | ||||| |   |  |  || |    |||||  | |||||
Db    576 TGGAGGGATAATGTCGAAAGCTTTCGAGTACATAATCAAAAACCAAGGCATCACTACTGA 635

Qy    662 TGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAA 721
          ||| | |||||  |||  ||    | |           || |   ||    |  |   |
Db    636 GGATAACTATCCATACCAAGAATCACA---------ACAAACTTGTAGTTCATCAACTAC 686

Qy    722 GGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGAC 781
            |  ||           || | |   ||| |    | ||||||        |    | |
Db    687 CCTATCATCGTCTTTTCGTGCAGCCACAATCAGCGGATATGAGACCGTTCCAATGAATAA 746

Qy    782 AGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAA 841
          ||| |||||||  ||   ||  | | || ||  ||  |||    || || |  |
Db    747 TGAGGAAGCGTTGCTACAAGCAGTGTCTCAACAGCCTGTTTCTGTGGGGATAGAAGGGAC 806

Qy    842 AGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGG 901
          ||        ||      |           |      |           |  || ||
Db    807 AGGGGCCGCGTTTAGGCACTACTCGGGAGGGGTATTCAACGGAGAATGTGGGACGGATTT 866

Qy    902 GATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGAT 961
          |  | ||    ||     |  |||    |          |  || ||     | || ||| |
Db    867 GCATCATGCGGTTACAATTGTTGGTTATGGAATGAGTGAAGAAGGGACTAAGTATTGGGT 926

Qy    962 AGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAA 1021
          | ||| ||||| ||||| ||| |||||||||| |||||||||  |||||  | ||| |
Db    927 GGTGAAGAATTCTTGGGGGGAAACTTGGGGAGAAAATGGTTACATGAGGATCAAGAGAGA 986

Qy   1022 CACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCA 1073
          |  || |        |||| |||||||  ||     | | ||||  || |||
Db    987 CGTGGATGCACCCCAAGGAATGTGTGGTTTGGCCATTCTTGCTTTCTATCCA 1038


RESULT 6
US-09-887-576-668
; Sequence 668, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.
;  APPLICANT:  Brown, D.
;  APPLICANT:  Chang, H.
;  APPLICANT:  Zhu, T.
;  APPLICANT:  Han, B.
;  APPLICANT:  Wang, X.
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: Promoters for regulation of plant expression
;  FILE REFERENCE: 1360.001US1
;  CURRENT APPLICATION NUMBER: US/09/887,576
;  CURRENT FILING DATE:  2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/213,848
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/214,087
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/258,692
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 875
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 668
;   LENGTH: 1047
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-887-576-668

  Query Match             8.9%;  Score 102.4;  DB 10;  Length 1047;
  Best Local Similarity   49.7%;  Pred. No. 4.9e-20;
  Matches  324;  Conservative    0;  Mismatches  316;  Indels   12;  Gaps    2;

Qy    425 GGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGG 484
          |||||||||  || | || |    | || |  || ||||| |||||  | ||||||  |  |
Db    396 GGATTGGAGACAAGAGGGGGCTGTTACACCCGTTAAGTATCAAGGCCGATGTGGAGGATG 455

Qy    485 TTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCT 544
          ||||||||||| ||   ||  ||  | ||||           ||  | | ||| || ||
Db    456 CTGGGCGTTTTCCGCGGTGGCGGCAGTGGAAGGCATTACAAAGATTACTAAAGGCGAGCT 515

Qy    545 TGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGCGAAGGTTCTTA 601
          |||    || || || ||| ||||| |||||||  |     ||  | |   ||||  | |
Db    516 TGTATCGCTATCGGAGCAACAACTCTTAGACTGCGACAGAGACTACAATCAAGGATGTCG 575

Qy    602 CAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGA 661
            ||      ||   || | ||||| |   |  |  || |    |||||  | |||||
Db    576 TGGAGGGATAATGTCGAAAGCTTTCGAGTACATAATCAAAAACCAAGGCATCACTACTGA 635

Qy    662 TGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAA 721
          ||| | |||||  |||  ||    | |           || |   ||    |  |   |
Db    636 GGATAACTATCCATACCAAGAATCACA---------ACAAACTTGTAGTTCATCAACTAC 686

Qy    722 GGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGAC 781
            |  ||           || | |   ||| |    | ||||||        |    | |
Db    687 CCTATCATCGTCTTTTCGTGCAGCCACAATCAGCGGATATGAGACCGTTCCAATGAATAA 746

Qy    782 AGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCAAA 841
          ||| |||||||  ||   ||  | | || ||  ||  |||    || || |  |
Db    747 TGAGGAAGCGTTGCTACAAGCAGTGTCTCAACAGCCTGTTTCTGTGGGGATAGAAGGGAC 806

Qy    842 AGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCGTATGG 901
          ||        ||      |           |      |           |  || ||
Db    807 AGGGGCCGCGTTTAGGCACTACTCGGGAGGGGTATTCAACGGAGAATGTGGGACGGATTT 866

Qy    902 GATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGAT 961
          |  | ||    ||     |  |||    |          |  || ||     | || ||| |
Db    867 GCATCATGCGGTTACAATTGTTGGTTATGGAATGAGTGAAGAAGGGACTAAGTATTGGGT 926

Qy    962 AGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAA 1021
          | ||| ||||| ||||| ||| |||||||||| |||||||||  |||||  | ||| |
Db    927 GGTGAAGAATTCTTGGGGGGAAACTTGGGGAGAAAATGGTTACATGAGGATCAAGAGAGA 986

Qy   1022 CACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCA 1073
          |  || |        |||| |||||||  ||     | | ||||  || |||
Db    987 CGTGGATGCACCCCAAGGAATGTGTGGTTTGGCCATTCTTGCTTTCTATCCA 1038


RESULT 7
```

US-09-770-445-20/c
; Sequence 20, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1282
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-20

Query Match 8.3%; Score 95.8; DB 10; Length 1282;
Best Local Similarity 52.2%; Pred. No. 4.7e-18;
Matches 319; Conservative 0; Mismatches 272; Indels 20; Gaps 4;

```
Qy    414 CCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGC 473
Db    862 CCAGACACAAAAGATTGGAGAGAAGATGGTATTGTTAGCCCTGTGAAAGAACAGGGACAT 803

Qy    474 TGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCA 533
Db    802 TGTGGATCTTGTTGGACATTTAGCACAACTGGAGCTCTTGAAGCAGCTTACCATCAAGCA 743

Qy    534 ACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGG------AAGAA 587
Db    742 TTTGGAAAAGGAATATCTTTGTCCGAGCAACAGCTTGTGGATTGTGCTGGTACTTTCAAT 683

Qy    588 AGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGT 647
Db    682 AACTTTGGTTGTCATGGTGGACTTCCTTCTCAAGCCTTTGAATACATTAAATACAACGGT 623

Qy    648 GGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAAT 707
Db    622 GGGCTCGACACGGAGGAGGCTTATCCTTACACCGGAAAAGACGGTGGCTGCAAATTTTCA 563

Qy    708 AAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGT 767
Db    562 GCGAAAAACATCGGTGTACAAGTCCGTGACTCTGTCAACATTACCCTGGGTGCAGAAGAT 503

Qy    768 ACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTC 827
Db    502 ---GAACTGAAGCACGCGGTCGGGTTGGTAAGGCCAGTGAGTG----------TGGCGTT 456

Qy    828 TCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACT-G 886
Db    455 TGAGGTTGTACATGAATTCAGGTTTTATAAGAAGGGAGTTTTTACTAGCAATACATGTGG 396

Qy    887 TACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGG 946
Db    395 TAACACTCCAATGGATGTAAACCATGCAGTGTTAGCAGTTGGTTATGGAGTTGAAGACGA 336

Qy    947 TGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACAT 1006
Db    335 CGTCCCGTACTGGCTTATAAAGAACTCATGGGGAGGTGAATGGGGAGACAATGGCTACTT 276

Qy   1007 TTGGATCCAAA 1017
Db    275 CAAGATGGAAA 265
```

RESULT 8
US-09-938-842A-2351
; Sequence 2351, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2351
; LENGTH: 1095
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2351

Query Match 7.5%; Score 86.8; DB 9; Length 1095;
Best Local Similarity 50.6%; Pred. No. 1.9e-15;
Matches 335; Conservative 0; Mismatches 297; Indels 30; Gaps 4;

```
Qy    426 GATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGT 485
Db    391 GATTGGCGAGAGAAAGGAGCTGTCACTGAGGTCAAGAATCAACAAGATTGTGGAAGTTGC 450

Qy    486 TGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTT 545
Db    451 TGGGCGTTTTCGACGGTTGCAGCAGTGGAAGGGATAAACAAGATCAGAACAAACAAACTG 510

Qy    546 GTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG---TGGAAGAAAGCGAAGGTTCTTAC 602
Db    511 GTTTCATTGTCTGAACAAGAGCTTGTGGATTGTGACACTGAAGAGAATCAAGGTTGTGCA 570

Qy    603 AATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTGGCATTGCCACTGAT 662
Db    571 GGAGGTCTCATGGAACCTGCGTTTGAATTTATAAAGAACAATGGTGGCATCAAAACCGAA 630

Qy    663 GATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAG 722
Db    631 GAGACTTATCCTTACGATTCCAGTGACGTTCAATTCTGTAGAGCTAATAGTATTGGTGGA 690

Qy    723 GTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACA 782
Db    691 GAAACTGTAACCATCGATGGACACGAACACGTCCCTGAGAATGAT---------------- 735

Qy    783 GAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGC---- 838
Db    736 GAGGAAGAACTTCTCAAAGCTGTTGCTCACCAGCCTGTCTCTGTAGCTATTGATGCTGGG 795
```

```
Qy    839 --AAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCG 896
          ||||||  ||  | ||| | | |||  | | |   ||||||  | | |       |
Db    796 AGCTCAGATTTCCAGCTTTACTCTGAGGGTGTGTTTATCGGAGAATGCGGGA------CT 849

Qy    897 TATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTAC 956
          |   ||   |     | ||  |  ||| |  |   |   |     ||||   | | ||
Db    850 CAGTTGAACCACGGGGTGGTGATTGTTGGGTATGGAGAGACCAAAAATGGAACAAAATAT 909

Qy    957 TGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAA 1016
          |||||||  |  || |||||||||   || ||||||||||  || ||  || ||||  ||
Db    910 TGGATAGTAAGGAACTCATGGGGACCTGAATGGGGAGAAGGAGGCTATGTTCGGATAGAA 969

Qy   1017 AGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACC 1076
          |||   |     |      ||||   || || ||          ||||| || || |||
Db    970 AGAGGAATATCGGAGAATGAAGGACGTTGCGGTATAGCCATGGAGGCTTCTTATCCCACC 1029

Qy   1077 AA 1078
          ||
Db   1030 AA 1031


RESULT 9
US-09-878-574-1099
; Sequence 1099, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 1099
;   LENGTH: 390
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: LIB3028-042-Q1-B1-B5
US-09-878-574-1099

  Query Match             7.4%;  Score 85.6;  DB 10;  Length 390;
  Best Local Similarity   58.1%;  Pred. No. 2.4e-15;
  Matches  173;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    1;

Qy    410 TCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGG 469
          ||  ||| || |    || ||||| ||| ||||| |  | |    ||| ||  | |||||
Db     78 TCTTCCTCCAACGAAAGACTGGAGAAAAGAAGGTATAGTGAGTTCAGTTAAAGATCAAGG 137

Qy    470 GGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAAT 529
            |||| |||    | ||| | ||     | || || ||  ||||||||||  |||||
Db    138 CAGCTGCGGATCATGCTGGACATTCAGCACAACTGGGGCTTTAGAAGCAGCCTATGCACA 197

Qy    530 AGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGG------A 583
          ||||   ||  |   | |    |||||||| ||  | || || |||||||  |
Db    198 AGCATTTGGGAAGAGTATCTCTCTTTCTGAGCAGCAGCTAGTGGACTGTGCTGGCCCTTT 257

Qy    584 AGAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACA 643
          | | |   ||  |   |   ||| | ||  |  ||| | || || |   ||  | |  |
Db    258 CAACAACTTTGGCTGCCATGGTGGGTTGCCATCACAAGCCTTTGAGTACATTAAATACAA 317

Qy    644 TGGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAA 701
          ||||||  | |  || || || |  ||||| |||| ||  ||||| |||   ||||||
Db    318 TGGTGGACTAGAGACAGAGGAAGCATATCCCTACACAGGAAAAGATGGTGTCTGCAAA 375


RESULT 10
US-09-770-149-933/c
; Sequence 933, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
;  APPLICANT:  Gorlach, Jorn
;  APPLICANT:  An, Yong-Qiang
;  APPLICANT:  Hamilton, Carol M.
;  APPLICANT:  Price, Jennifer L.
;  APPLICANT:  Raines, Tracy M.
;  APPLICANT:  Yu, Yang
;  APPLICANT:  Rameaka, Joshua G.
;  APPLICANT:  Page, Amy
;  APPLICANT:  Matthew, Abraham V.
;  APPLICANT:  Ledford, Brooke L.
;  APPLICANT:  Woessner, Jeffrey P.
;  APPLICANT:  Haas, William David
;  APPLICANT:  Garcia, Carlos A.
;  APPLICANT:  Kricker, Maja
;  APPLICANT:  Slader, Ted
;  APPLICANT:  Davis, Keith R.
;  APPLICANT:  Allen, Keith
;  APPLICANT:  Hoffman, Neil
;  APPLICANT:  Hurban, Patrick
;  TITLE OF INVENTION:  Expressed Sequences of Arabidopsis
;  TITLE OF INVENTION:  thaliana
;  FILE REFERENCE:  2024 (PARA-013PRV)
;  CURRENT APPLICATION NUMBER: US/09/770,149
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER:  60/178,506
;  PRIOR FILING DATE:  2000-01-27
;  NUMBER OF SEQ ID NOS:  999
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 933
;   LENGTH: 604
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-770-149-933

  Query Match             7.0%;  Score 80.8;  DB 10;  Length 604;
  Best Local Similarity   58.5%;  Pred. No. 7.6e-14;
  Matches  186;  Conservative    0;  Mismatches  117;  Indels   15;  Gaps    2;

Qy    783 GAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCAATTGATGCA--- 839
          |||  |     |||| |  ||  |   | | || ||  | ||||||  | |||||||||
Db    473 GAGAAGAGTCTCTTGAAGGCATTGGCTCATCAGCCTCTCAGTGTCGCCATTGATGCATCT 414

Qy    840 ---AAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAACTGTACAAGTCCG 896
              | ||| || || || || | ||| ||  | | ||||||      | |        |
Db    413 GGTAGAGAGTTCCAGTTCTATAGCGGCGGCGTGTTTGATGGGCGGTGCGG---------G 363

Qy    897 TATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTAC 956
            ||   || | |||  |||       || || ||||| |||    | |||  |||||||
Db    362 GTTGATCTTGACCACGGTGTGGCTGCGGTTGGGTATGGATCAAGCAAGGGTTCAGATTAC 303

Qy    957 TGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAA 1016
            || | ||| ||||| ||||||  | | |||||||||  | ||||||||  || |  |
Db    302 ATCATTGTGAAGAATTCTTGGGGACCAAAATGGGGAGAAAAAGGTTACATCAGGCTGAAG 243

Qy   1017 AGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGCTTCATACCCAACC 1076
          || ||||| || ||   |   ||  | ||||| || ||  |  | ||||| | ||| |||
Db    242 AGGAACACTGGGAAACCAGAGGGTCTCTGTGGAATCAACAAGATGGCTTCTTTCCCCACC 183

Qy   1077 AAAGAGGAATCAGAAACA 1094
          |||   | | |  | ||
Db    182 AAAACTAAGTGATCATCA 165


RESULT 11
US-09-938-842A-1324
; Sequence 1324, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
```

```
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1324
;   LENGTH: 1122
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-1324

  Query Match             6.4%;  Score 74;  DB 9;  Length 1122;
  Best Local Similarity   62.4%;  Pred. No. 1.1e-11;
  Matches  116;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    407 CCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCA 466
          | |||  ||  |    |  |||||| |  || |||| | | |||| |  || ||  ||||
Db    414 CGATCTTCCGACGGAGTTCGATTGGCGTGAACAAGGAGCCGTCACACCTGTTAAAAACCA 473

Qy    467 AGGGGGCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGC 526
            ||    |||||     | ||| | |||  ||| |  |||||  | |||| |||||||
Db    474 GGGTATGTGTGGTTCATGCTGGTCATTTAGTGCGATAGGAGCTCTTGAAGGAGCACATTT 533

Qy    527 AATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGA 586
            |||| ||   ||| ||||||||||   ||| ||  | |||||||| ||||   | ||
Db    534 TCTAGCCACTAAAGAGCTTGTTAGCCTCAGTGAGCAGCAGCTCGTAGATTGTGACCATGA 593

Qy    587 AAGCGA 592
            | ||
Db    594 GTGTGA 599


RESULT 12
US-09-878-574-10000
; Sequence 10000, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 10000
;   LENGTH: 264
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 701103159H1
US-09-878-574-10000

  Query Match             6.3%;  Score 72.4;  DB 10;  Length 264;
  Best Local Similarity   66.9%;  Pred. No. 1.4e-11;
  Matches  103;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy    426 GATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGGGGT 485
          ||||||||  | |  || |   | || |  || ||  ||||||   |||||||     ||
Db     64 GATTGGAGACAGAGCGGCGCAGTTACACCCGTCAAAGACCAAGCAAGCTGTGGTGCATGT 123

Qy    486 TGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTT 545
          |||||||| || || || || || || |||    | ||   || |  |||||    |||
Db    124 TGGGCGTTCTCAGCTACTGGTGCTATTGAAGGCATAAATAAGATTGTCACAGGGTCTCTT 183

Qy    546 GTTAGCCTTTCTGAACAAGAACTCGTAGACTGTG 579
          || |||||||||||||||| ||  |||| ||||
Db    184 GTAAGCCTTTCTGAACAAGAGCTGATAGATTGTG 217


RESULT 13
US-10-099-275-1
; Sequence 1, Application US/10099275
; Publication No. US20020187499A1
; GENERAL INFORMATION:
;  APPLICANT: Schneider, Patrick
;  APPLICANT:  Yamamoto, Karen K.
;  APPLICANT:  French, Cynthia K.
;  APPLICANT:  Reprogen, Inc.
;  TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of
;  TITLE OF INVENTION:  Endometriosis
;  FILE REFERENCE: 018002-001310US
;  CURRENT APPLICATION NUMBER: US/10/099,275
;  CURRENT FILING DATE:  2002-04-13
;  PRIOR APPLICATION NUMBER: US/09/701,685
;  PRIOR FILING DATE: 2001-03-26
;  PRIOR APPLICATION NUMBER: US 60/088,017
;  PRIOR FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1643
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (70)..(1065)
;   OTHER INFORMATION: human cathepsin S
US-10-099-275-1

  Query Match             6.2%;  Score 72.2;  DB 9;  Length 1643;
  Best Local Similarity   49.2%;  Pred. No. 4.4e-11;
  Matches  324;  Conservative    0;  Mismatches  308;  Indels   27;  Gaps    4;

Qy    414 CCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGC 473
          ||||  ||   ||| |||||  | |||||     | ||  |||| || || |||||
Db    415 CCTGATTCTGTGGACTGGAGAGAGAAAGGGTGTGTTACTGAAGTGAAATATCAAGGTTCT 474

Qy    474 TGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCA 533
          |||||     | ||||| ||   |||   ||| ||| | ||||||   |      |   |
Db    475 TGTGGTGCTTGCTGGGCTTTCAGTGCTGTGGGGGCCCTGGAAGCACAGCTGAAGCTGAAA 534

Qy    534 ACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGA------- 586
          |||||| | || ||    ||   ||  ||  | || || || ||       ||
Db    535 ACAGGAAAGCTGGTGTCTCTCAGTGCCCAGAACCTGGTGGATTGCTCAACTGAAAAATAT 594

Qy    587 --AAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACAT 644
            || | |||| |   |   ||| |    |       | ||| | |   |  | ||  |
Db    595 GGAAACAAAGGCTGCAATGGTGGCTTCATGACAACGGCTTTCCAGTACATCATTGATAAC 654

Qy    645 GGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCC 704
            ||||| | | | || | |   ||||| |||| ||| |      | |  || || | |
Db    655 AAGGGCATCGACTCAGACGCTTCCTATCCCTACAAAGCCA-----TGGATCTGAAATGTC 709

Qy    705 AATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAG 764
          |||| ||  ||| | |  ||        | |   |    || || | | |  |  |||
Db    710 AATATGACTCAAAATATCGTGCTGCCACATGTTCAAAGTACACTGA-ACTTCCTTATGGC 768

Qy    765 AGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGT 824
          ||   |||       || |||    | |  |   |  | |           ||| |   |
Db    769 AGAGAAGATGTCCTGAAAGAAGCTGTGGCCAATAAAGGC------------CCAGTGTCT 816

Qy    825 GTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAAC 884
          ||     | |||||     ||  |  ||| | |  |    ||| |  ||      |  |
Db    817 GTTGGTGTAGATGCGCGTCATCCTTCTTTCTTCCTCTACAGAAGTGGTGTCTACTATGAA 876

Qy    885 TGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGAT 944
```

```
Db    877 CCATCCTGTACTCAGAATGTGAATCATGGTGTACTTGTGGTTGGCTATGGTGATCTTAAT 936

Qy    945 GGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTAC 1004
          ||   ||| |||||| | | ||||||    |||||  |  | |  || ||||| || ||
Db    937 GGGAAAGAATACTGGCTTGTGAAAAACAGCTGGGGCCACAACTTTGGTGAAGAAGGATAT 996

Qy   1005 ATTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTATTTCGC 1063
          ||| ||||   |||||| |  || |||   |  ||  |   | |  |    | || | |
Db    997 ATTCGGATGGCAAGAAATAAAGGAAATCATTGTGGGATTGCTAGCTTTCCCTCTTACCC 1055


RESULT 14
US-09-887-576-294
; Sequence 294, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.
;  APPLICANT:  Brown, D.
;  APPLICANT:  Chang, H.
;  APPLICANT:  Zhu, T.
;  APPLICANT:  Han, B.
;  APPLICANT:  Wang, X.
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: Promoters for regulation of plant expression
;  FILE REFERENCE: 1360.001US1
;  CURRENT APPLICATION NUMBER: US/09/887,576
;  CURRENT FILING DATE:  2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/213,848
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/214,087
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/258,692
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 875
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 294
;   LENGTH: 2004
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-887-576-294

  Query Match             6.2%;  Score 71.2;  DB 10;  Length 2004;
  Best Local Similarity   58.6%;  Pred. No. 9.6e-11;
  Matches  143;  Conservative    0;  Mismatches   98;  Indels    3;  Gaps    1;

Qy    472 GCTGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAG 531
          | |  || ||  ||||||| || ||  | |  |||||  | |||| |  | |    || |
Db   1251 GTTTAGGGAGTTGTTGGGCATTCTCAACTATTGGAGCAGTGGAAGGAATAAACAAGATTG 1310

Qy    532 CAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAG---AAA 588
            ||||||||  |  |     | ||||||||||||| | || || ||||    |    | |
Db   1311 TGACAGGAGATTTAATATCTTTGTCTGAACAAGAATTGGTTGATTGTGACACATCTTATA 1370

Qy    589 GCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACATGGTG 648
          | |||| | | |   ||| |    | |  |  |||| || |   || |  || |||||||
Db   1371 ACCAAGGATGTAATGGTGGTTTAATGGACTATGCGTTTGAGTTTATTATCAAAAATGGTG 1430

Qy    649 GCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATA 708
          | ||||  || || |  |||||||||||||  |||   || ||| | ||  |    || |
Db   1431 GTATTGATACCGAAGCGGATTATCCTTACAAGGCTGCTGATGGTCGTTGTGACCAGAACA 1490

Qy    709 AGAT 712
            | |
Db   1491 GGGT 1494


RESULT 15
US-09-990-064-1
; Sequence 1, Application US/09990064
; Patent No. US20020164765A1
; GENERAL INFORMATION:
;  APPLICANT: Parvinder K. Punia
;  TITLE OF INVENTION: NOVEL COMPOUNDS
;  FILE REFERENCE: GP-30242
;  CURRENT APPLICATION NUMBER: US/09/990,064
;  CURRENT FILING DATE:  2001-11-21
;  PRIOR APPLICATION NUMBER: GB0028462.0
;  PRIOR FILING DATE: 2000-11-22
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 996
;   TYPE: DNA
;   ORGANISM: Mammalia
US-09-990-064-1

  Query Match             5.9%;  Score 68.2;  DB 9;  Length 996;
  Best Local Similarity   47.2%;  Pred. No. 4.9e-10;
  Matches  300;  Conservative    0;  Mismatches  308;  Indels   27;  Gaps    2;

Qy    414 CCTGCATCATGGGATTGGAGGAAAAAAGGTGTCATCACCCAAGTAAAGTACCAAGGGGGC 473
          ||||  ||   ||| ||||| | |  ||     ||||  |||| || || || ||
Db    346 CCTGATTCTGTGGACTGGAGAGAGAGGGGGTGTGTCACTGAAGTGAAATATCAGGGTTCT 405

Qy    474 TGTGGAAGGGGTTGGGCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCA 533
          |||||     | ||||| ||   |||| ||| ||  | ||||||   |     |
Db    406 TGTGGTGCTTGCTGGGCCTTCAGTGCCGTGGGGGCTCTGGAAGCACAGCTGAAGCTGAAG 465

Qy    534 ACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGT---------GGAA 584
          |||||| | || ||    ||   || |||  | || || || ||            | |
Db    466 ACAGGAAATCTGGTGTCTCTCAGTGCACAGAATCTGGTGGATTGCTCAACTACAAAGTAC 525

Qy    585 GAAAGCGAAGGTTCTTACAATGGATGGCAGTATCAATCGTTCGAATGGGTTTTAGAACAT 644
          |  | | |||| |   |   ||| |    |    | ||||| |||   |  | ||  |
Db    526 GGGAACAAAGGCTGCAATGGTGGCTTCATGACAGAGGCGTTCCAATACATCATCGATAAC 585

Qy    645 GGTGGCATTGCCACTGATGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCC 704
            ||||| |   | || | |   |||| |||| ||| |  ||    |  |||. |
Db    586 AACGGCATCGATTCAGAAGCTTCCTATCCCTACAAAGCCATGGATCAAAAGTGCCATT-- 643

Qy    705 AATAAGATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAG 764
                      | |   |||   |  |   ||  ||      || ||     ||
Db    644 ----------------ATGACTCAAAACATCGTGCTGCCACCTGTTCAAAGTACACAGAA 687

Qy    765 AGTACAGAATCAGAGACAGAGCAAGCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGT 824
            | |             ||||     |      |   |     |  | |   |||  |   |
Db    688 CTTCCCTTTGGCAGTGAAGAGGCCTTGAAAGAAGCTGTGGCCAATAAAGGACCTGTGTCT 747

Qy    825 GTCTCAATTGATGCAAAAGATTTTCATTTATACACCGGGGGAATTTATGATGGAGAAAAC 884
          ||  | || |||||||   ||| |  ||| | |  |    ||| |  ||       |  |
Db    748 GTGGCTATAGATGCAAGTCATTCTTCTTTCTTCCTCTACAGAAGTGGTGTCTATTATGAA 807

Qy    885 TGTACAAGTCCGTATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGAT 944
            |  || |  |      | |||||    ||  ||   || || ||||||       |
Db    808 CCATCCTGTACTCAGAATGTGAATCATGGCGTACTAGCAGTTGGCTATGGTAACCTTAAG 867

Qy    945 GGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGATGGTTAC 1004
          ||   ||| |||||| | | ||||||    |||||     | |  || ||| | || ||
Db    868 GGGAAAGACTACTGGCTTGTGAAAAACAGCTGGGGCATCCACTTCGGTGAACAAGGATAT 927

Qy   1005 ATTTGGATCCAAAGAAACACGGGTAATTTATTAGG 1039
          ||| ||||   ||||||||      |||   |  ||
Db    928 ATTCGGATGGCAAGAAACAGTAAAAATCATTGTGG 962


Search completed: January 11, 2003, 22:08:22
Job time : 70 secs
```